DOCUMENT RESUME

ED 043 566 SP 004 147

AUTHOR	Klausmeier, Herbert J., Ed.; O'Hearn, George T., Ed.
TITLE	Research and Development Toward the Improvement of Education.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
PUB DATE	68
CONTRACT	OEC-5-10-154
NOTE	171p.
AVAILABLE FROM	Dember Educational Research Services, Inc., Box 1148, Madison, Wis. 53701 ($5.75)
EDRS PRICE	EDRS Price MF-$0.75 HC-$8.65
DESCRIPTORS	*Curriculum Research, *Educational Development, *Educational Research, English Education, *Instructional Media, Interaction Process Analysis, *Learning Processes, Mathematics Education, Microteaching, Models, Research and Development Centers, Science Education

ABSTRACT

This book contains 17 papers prepared for a colloquium series on development and related research dealing with conditions and processes of learning, subject matter content and sequence, instructional media, and instructional methods and teacher behaviors: 1) "Organized Retention and Cued Recall" by Endel Tulving; 2) "Developmental Processes in Thought," Kenneth Lovell; 3) "Promoting Creative Thinking in the Classroom," Martin Covington; 4) "Measuring Motivation in Culturally Disadvantaged School Children," Richard deCharms; 5) "Curriculum Research in Mathematics," E.G. Besle; 6) "Curriculum Research in Science," Arthur H. Livermore; 7) "Curriculum Research and Development in English," Robert Pooley; 8) "Research in Reading," Wayne Otto; 9) "Strategies for Concept Attainment in Mathematics," Myron Rosskopf; 10) "Instructional Television Around the World," Wilbur Schramm; 11) "Instructional Research: Some Aspects of Its Status, Defects, and Needs," Arthur Lumsdaine; 12) "SOCRATES, A Computer-Based Instructional System in Theory and Research," Lawrence Stolurow; 13) "An Analytical Approach to Research on Instructional Methods," Nathanial Gage; 14) "Interaction Analysis and Inservice Training," Ned Flanders; 15) "Research and Development Strategies: The Current Scene," Louis Bright; 16) "The Wisconsin Research and Development Center for Cognitive Learning," Herbert Klausmeier; 17) "An Output-Oriented Model of Research and Development and Its Relationship to Educational Improvement," Hendrick Gideonse. (JS)

$5.75

IN THIS BOOK 20 eminent scholars, researchers, and research administrators present their latest thinking about the processes and conditions of human learning, the processes and programs of instruction, and the application of research and development strategies to educational problems.

As the editors point out in the Preface, the application of research and development strategies for the improvement of educational practices is a pioneering venture begun in 1964 under the provisions of the Cooperative Research Program of the United States Office of Education and now continuing through Title IV of the Elementary and Secondary Education Act of 1965.

Working from a proposed model outlined in Chapter 16, the editors have organized the book around the major components of an instructional system.

Basic to the improvement of education is knowledge about the processes and conditions necessary for efficient learning. In Part I, four authors discuss retention and recall, thought processes, creative thinking, and motivation.

A major component of an instructional system is subject matter content and sequence. In Part II, five authors provide scholarly insights into the structures of such disciplines as mathematics, science, English, and reading. They state or imply that the structure of the discipline itself, as formulated by one or more scholars, has provided much of the basis for recent curriculum development.

Instructional materials and media can be considered as the interface between the learner and the subject matter being learned. In Part III, the discussion focuses on instructional television, on some aspects of the status, defects, and needs of instructional research, and on computer-based instruction.

Teaching methods, or the interactions which occur between a teacher and students, have been the subject of many research studies in the last decades. In Part IV, two recent paradigms for research on teaching are presented. One approach discusses what is commonly called micro-teaching and the other is interaction analysis.

Finally, in Part V, three research administrators discuss the current application of research and development strategies for the improvement of educational practices. The present scene in educational R & D is one of rapid change brought about largely by the advent of Federal funding. The development of these changes with reference to Federal involvement is traced in one chapter. In another chapter the focus and programs of the Wisconsin Research and Development Center for Cognitive Learning are described. And in the final chapter an author presents an output-oriented model of R & D.

Research and Development Toward the Improvement of Education grows out of a colloquium series sponsored recently by the Wisconsin R & D Center. The book has been published by Dembar Educational Research Services, Inc., as a contribution to the field, without financial support other than that engendered by sales.

T [redacted] **Research and Development Toward the Improvement of Education**[1]

Edited by

A [redacted] Herbert J. Klausmeier and George T. O'Hearn

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

inst. source

DEMBAR EDUCATIONAL RESEARCH SERVICES, INC. Box 1148 Madison, Wisconsin 53701
Publisher

Research and Development Toward the Improvement of Education

Published by Dembar Educational Research Services, Inc., Madison, Wisconsin.
Printed in the United States of America.

First Edition.

Library of Congress Catalog Card Number 68-58196

[1] The preparation of this report was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program.

Center No. C-03/Contract OE 5-10-154

PREFACE

THE FIRST systematic attempt to apply research and development strategies to education occurred with the establishment of four research and development (R & D) centers in 1964 as part of the Cooperative Research Program of the United States Office of Education. The R & D Centers program attempted to combine the best features of the separate programs of the U.S.O.E.: basic and applied research, development, curriculum improvement, and demonstration. Thus, an R & D Center is organized to conduct basic research related to a significant problem area; to develop substantive and procedural products and conduct related research; to demonstrate the results of the research and development activities in a few school settings; and to take initiative for the eventual widespread dissemination of the research results and tested products. An R & D Center, located on a university campus, draws its professorial staff from the behavioral sciences, subject disciplines, communications, educational methodology, and other fields. It is a community of scholars, capable of dealing with a problem area effectively. In an R&D Center, then, human and financial resources are concentrated on a particular problem area over an extended period of time in order to secure a better understanding of that problem area and to improve related educational practices. Since 1964 the number of R & D Centers has increased from four to nine.

Our major universities, including the nine where R & D Centers are located, have a large pool of able researchers and many other resources. However, the development and application of research and development strategies to education is a pioneer venture. Developing imaginative strategies for extending knowledge about cognitive learning and improving related educational practices is a continuing concern of the Wisconsin Research and Development Center for Cognitive Learning.

In this Center, basic research is done to extend knowledge about the processes of cognitive learning, particularly those associated with concept learning and problem solving, and about the conditions associated with efficiency of learning in the cognitive domain. Some of this knowledge goes immediately into the development of instructional systems. A second objective of the Wiscon-

sin R & D Center is the improvement of educational practices in the cognitive domain. The development and refinement of an instructional system related to a subject field, such as mathematics, requires appropriate strategies for dealing with the various components independently and also for interrelating them so that desired behavioral objectives are achieved. Three important components of the system are subject matter content and sequence; instructional materials and media; and instructional methods, including teacher behaviors and learner activities. A colloquium series was organized to bring information to the Center staff regarding these research and development strategies.

One part of this colloquium series was concerned with research on learning that has fairly direct implications for improving education. (A previous volume, Analyses of Concept Learning, Academic Press, 1966, on the other hand outlined more basic research.) The main portion of the colloquium series was addressed to development and related research dealing with subject matter content and sequence, instructional media, and instructional methods. This volume is organized in the same pattern.

As might be expected, the research methods reported by the participants are varied. In Part I, Endel Tulving of the University of Toronto shows that substantive results concerning information storage and retrieval, derived from the typical controlled laboratory experiment, require us to reconsider our notions about why persons do not effectively reproduce information they once learned. Kenneth Lovell, University of Leeds, shows that the charting of thought processes developmentally is moving ahead rapidly by means of Piaget's methods and refined variants thereof. Martin Covington, University of California, describes a straightforward attempt to develop materials designed to facilitate productive thinking in school children and to utilize the material in extending knowledge about children's productive thinking skills. Here there is a continuous interweaving of theorizing, developing material based on theory, and research activities concerning the abilities nurtured by the instructional program. Richard DeCharms and Virginia Carpenter of Washington University and Lindenwood College proceed in a similiar manner as Covington; however, their main concern is with motivational theory and the measurement of strength of motivation in culturally disadvantaged children. In these chapters substantive research results and theory properly receive most attention; however, sufficient attention is given to the procedures to make clear that extending knowledge about processes and conditions of learning can be accomplished by a variety of strategies.

In Part II, somewhat similar curriculum development and research procedures are outlined by the first three authors. E. G. Begle, Stanford University, outlines development and related research in mathematics, including his experiences with the widely known School Mathematics Study Group. Arthur Livermore, AAAS Commission on Science Education, briefly describes the large science cur-

riculum projects, kindergarten to Grade 12, but gives major attention to the elementary program: Science—A Process Approach. Robert C. Pooley, The University of Wisconsin, also synthesizes many research and development activities related to the teaching of literature, composition, and language but gives particular attention to the Wisconsin English–Language Arts Project. Common to the curriculum development strategies in these subjects is an increasing emphasis on the structure of the discipline and a decreasing emphasis on other criteria such as social utility, difficulty, children's interests, and possible transfer to other subject fields. In the last two chapters of Part II, Wayne Otto, The University of Wisconsin, describes a rigorous and productive research effort in reading, and Myron Rosskopf, Teachers College, Columbia University, relates curriculum development in mathematics to recent research in concept learning.

In Part III, Wilbur Schramm, Stanford University, reports information on the use of instructional television. Three case studies are presented in sufficient detail to renew confidence in the value of careful observation in the natural environment by an interested, objective scholar. Lawrence Stolurow, Harvard University, describes the computer-based instructional system, SOCRATES, that he developed while at the University of Illinois. Problems associated with development of the hardware and software are noted briefly or may be inferred. The use of the computer in research to improve instruction and to validate behavior theory is noteworthy. Arthur Lumsdaine, University of Washington, deals with research and development in the entire field of instructional media.

In Part IV, N. L. Gage, Stanford University, outlines a recent concept, micro-teaching, and related research strategies. The concept promises to generate much research and development on instructional methods. In [illegible] approach, the teacher and student rather than the content, hardware, or program, are central in the educative process. Ned Flanders, University of Michigan, describes the new directions of his long-standing research on teacher-student interaction analysis. In the current model, the number of categories of interactions is reduced. By employing a computer for rapid feedback, the current model holds much promise of providing immediate (daily) correction of teacher–student interactions and also for supplying the extensive research data required in validating and extending the theoretical basis of the system of interaction analysis.

Part V of this volume contains papers which did not arise from the colloquium series. They are included because they contribute to the survey of research and development activities in education, particularly very recent trends. Louis Bright and Hendrick Gideonse of the U.S.O.E. outline the broad sweep of research and development strategies and activities, resulting from the passage of the Elementary and Secondary Education Act of 1965. Included here are the R & D centers, regional educational laboratories, Title III centers, and other R & D instrumentalities funded by other governmental agencies. The comprehensive edu-

cational research and development program of the U.S.O.E. merits careful attention by anyone interested in the improvement of education. A more complete description of one federally-supported R & D instrumentality, the Wisconsin Research and Development Center for Cognitive Learning, follows the chapter by Bright and Gideonse. Here we review the main components of a Center and their interrelations. Particular attention is given to the problem area of the Center; its programs, strategies, and outcomes; staffing pattern; administrative organization; locus of operation; and source and stability of funding.

In a new field, models are required as guides to operations. Many models have been outlined to indicate how systematic improvement in education might be managed through sequential research, development, and dissemination. None of the models thus far appeared particularly appropriate to the current scene. Hendrick Gideonse, active privately and not as a U.S.O.E. official, has been attempting for some time to formulate a comprehensive model that would simultaneously meet the test of reality and also provide a more adequate conceptualization of Research and Development activities designed to improve education. Fortunately his conceptualization was completed at the time the volume was in final editing. The Gideonse model is especially relevant to those who have been concerned with relating research, development, and educational improvement.

The colloquium series was planned and executed by the editors of this volume; Herbert J. Klausmeier assumed primary responsibility for organizing the colloquium series around topical areas of relevance to the R & D Center. He identified most of the participants and the broad organization framework for each participant's contribution. George T. O'Hearn participated in the planning and organization of the colloquium series, worked out the presentation schedule and other details with each participant, and handled editorial queries from the publisher. He also edited the papers and wrote the introductions to each part. Barbara Kennedy, Project Assistant, read the papers, checked the references, and put the references in APA style. The Wisconsin Research and Development Center for Cognitive Learning, the education research and development community, and the teaching profession generally is indebted to the internationally recognized scholars who made this volume possible.

Herbert J. Klausmeier, Professor of
Educational Psychology; Director,
Wisconsin Research and Development
Center for Cognitive Learning

George T. O'Hearn, Associate Professor,
Acting Director, Division of Education
University of Wisconsin–Green Bay

September 1968

CONTENTS

PART I

Conditions and Processes of Learning

UNDERSTANDING learning phenomena and conditions associated with efficiency of learning is a foremost concern of psychologists throughout Western civilization. Increasingly sophisticated experiments are being executed in laboratory and non-laboratory settings. Representative of work in this vast domain, with implications for education, are the following:

a) controlled observations to infer thought processes.
b) controlled experimentation in the laboratory to extend our knowledge of human memory.
c) development of material and related research to nurture the creative abilities of school children.
d) development of dependent measures to extend a motivational theory to school settings.

Permanence of learning necessarily requires the storage and recall of information. In Chapter 1, Endel Tulving discusses various verbal learning and memory experiments conducted in the laboratory at the University of Toronto, and suggests that "the success of retention and recall is to a large extent determined by what the subject does with the material he is to memorize, by the methods he uses to organize the material, and by the cues that he uses to effect the recall of the material." The data reported by Tulving support his contention that the ability to recall is determined more by the limitation of the retrieval mechanism than by the limitation of the storage mechanism. He concludes:

> *. . . that understanding of many phenomena of verbal learning and memory would be enhanced if we adopted a basic orienting attitude. That a person's mnemonic performance depends to a very large extent on the method that he wittingly or unwittingly uses to organize the material. The important function of such organization is to facilitate retrieval of information from the memory storage.*

The inferring of thought processes has always been difficult but is essential to explaining changes that occur ontogenetically. In "Developmental Processes in Thought," Kenneth Lovell, University of Leeds, reports on the stages of

human intellectual growth hypothesized by Jean Piaget and his followers. Lovell suggests that the schemas interiorized by the child undergo constant modification and are the basis and limitation of all subsequent learning. "The implication of Piaget's findings is that we need to begin to stimulate the intellectual growth of the child in the early weeks of life." Research reported by Lovell confirms the general classification of the developmental stages of thought hypothesized by Piaget, and also indicates the importance of previous learning and the cultural milieu on the child.

Martin Covington in his paper "Promoting Creative Thinking in the Classroom" states, "In order to take full advantage of the unrestrictive atmosphere the child must first come to understand what constitutes creative ideas in the given situation, and how he can achieve such ideas for himself . . . In order to think creatively the child must have at his disposal a repertoire of creative thinking skills. To name a few: the ability to recognize gaps in existing information; a facility for formulating relevant questions, and a facility to the demands of the task so one can adequately judge the suitability of the proposed ideas."

The General Problem Solving Program described by Covington represents an attempt to initiate creative thought in students within a school setting and within a curriculum framework. The objective is to teach a number of cognitive skills which are fundamental to creative thinking and then determine whether such generalized skills can be applied to a specific subject matter area.

Before specified learning can take place the learner must seek and accept the learning situation. The probability that an individual will seek or accept a new learning situation is referred to as motivation by Richard deCharms and Virginia Carpenter. When a tendency to avoid a learning situation increases, i.e.., when there is a lack of motivation, the problem becomes crucial. The discussion of this problem by deCharms and Carpenter is based on their work with inner-city youngsters. In their chapter "Measuring Motivation in Culturally Disadvantaged School Children" the authors propose a model of motivation in which a person perceived to be originating his own behavior is called an "Origin" while a person perceived to be behaving in a certain way because of an external "power source" is said to be acting like a "Pawn." "Pawn behavior implies powerlessness, but origin behavior implies freedom of choice, not necessarily the ability to wield power over others." The theoretical framework of the "OP" model is discussed in detail. The n achievement procedures developed by McClelland were used in measuring motivation in a sample of culturally disadvantaged school children. The data obtained tend to support the theoretical point of view expressed.

Organized Retention and Cued Recall[1]

Endel Tulving
University of Toronto
Toronto, Ontario

EVER SINCE Hermann Ebbinghaus (1885) brought phenomena of memory under experimental scrutiny in the psychological laboratory and William James (1890) eloquently epitomized the distilled wisdom of other great thinkers on the subject, research on remembering and forgetting has been growing by leaps and bounds. Everyone agrees that this research has by now produced an immense quantity of empirical data, but few are willing to claim that it has created any genuine novel conceptual insights. More often than not, new empirical findings have generated more heat than light at the theoretical level. Thus, despite everything we believe we know about the role of a multitude of experimentally manipulable variables on verbal learning, retention, and recall, we seem to be no nearer to the understanding of the basic nature of memory than were psychologists and other writers of previous generations. If someone asked us today what do we know that is truly worth knowing and that was not known to Ebbinghaus and James, we would be hard pressed to provide a convincing answer.

Those who share this pessimistic assessment of the state of affairs in research on verbal learning and memory have been quick to suggest remedies, but their advice has been inconsistent. Some have argued that experimental conditions between laboratories be more rigidly standardized, others would like to see a greater diversity of methods. Some advise that only limited sets of carefully calibrated materials be used, others insist on a fuller variety of materials. Some are convinced that true understanding of memory can be achieved only through mammoth parametric studies, others are equally certain that at least for the time being the solution lies in simpler "miniature" experiments. I am listing these conflicting attitudes not to pass judgment on them, but rather as a background for my own favorite complaint about the general orientation of research in this broad field.

My complaint is that far too little attention has been devoted to the implications of the well-known fact that subjects in verbal learning and memory experiments are not passive organisms. Too many students of human learning and memory have predicated the interpretation of their data and the formulation of their theoretical statements on the basic assumption that in learning and remembering certain things happen to the person or certain events take place in his nervous system, and that these events are governed by nothing more than the present and past inputs. Eventually, of course, science may be able to interpret all phenomena of memory without recourse to a highly sophisticated, conscious, active, thinking organism, but at the present time nothing much seems to be gained, and a good deal may be lost, if we pretend that the articulate human subject is a passive recipient of external stimulation, and that acquisition, retention, recall, transfer, interference, etc., occur in a mental vacuum. If we know anything at all about what subjects do in verbal learning and memory experiments, we know that they use plans (Miller, Galanter, and Pribram, 1960), that they consciously and deliberately employ strategies (e.g., Pollack, Johnson, and Knaff, 1959) or mediational devices (e.g., Bugelski, 1962; Underwood and Schulz, 1960), or simply that they think (e.g., Mandler, 1965). There are some who would

argue that these covert activities of the subject, and their behavioral manifestations, merely accompany processes of memory and that they play no significant role in determining the quantity and quality of a person's mnemonic performance, but such views are rapidly losing their appeal. A great deal of evidence accumulated in recent years, added to evidence resurrected from isolated and forgotten experiments from the past, points to the witting or unwitting use of plans, strategies, and mediational devices by the subject as the sine qua non of remembering. A fuller understanding of what the subject does when faced with the requirements of memory tasks of various kinds, therefore, seems indispensable for the comprehension of many, if not all, phenomena of memory. In all probability, an active search for such understanding will turn out to be more fruitful than the mere introduction of standardized conditions of experimentation, calibration of materials, or mammoth experiments.

In the present paper I will describe some recent work we have been doing at Toronto to probe into some of the long-standing problems of memory. Two questions that I have been particularly interested in are, first why is repetition effective in increasing the amount of material the subject can recall, and secondly, how does retrieval of mnemonic information from the memory storage take place. I will obviously not be able to answer these questions, but I would like to offer some suggestions relevant to them. These suggestions are firmly anchored to the basic assumption that the success of retention and recall is to a large extent determined by what the subject does with the material he is to memorize, by the methods he uses to organize the material, and by the cues that he uses to effect the recall of retained material. The data I will report, I believe, will offer good support to this assumption.

FREE RECALL AND ORGANIZATION

The experimental data come from experiments in which the subject's task is to learn, that is to commit to memory, a set of discrete verbal items. Some of these experiments are concerned with free recall of items, while others constitute variations on the theme of free recall. In the free-recall experiment the subject is presented with a list of familiar items, such as meaningful words, and his task is to remember as many of these items as possible, in any order that they occur to him. Thus the free-recall paradigm constitutes a test for the acquisiton, retention, and utilization of retained information about the membership of otherwise familiar items in an unfamiliar set. Free recall as a laboratory task, like other paradigms of verbal learning and memory, involves mnemonic performance of the kind often required of individuals outside the laboratory. A teacher's knowledge of names of students in a class, a person's memory for the names of guests at a cocktail party he attended the night before, and a child's recollection of animals he saw on his visit to the zoo are based on retention of information of the kind that we investigate in free-recall experiments.

I have been particularly interested in free-recall phenomena for three reasons. First, the requirements imposed on the mnemonic information processing mechanisms in the free-recall task are necessarily present in any other paradigm. The subject cannot learn the order of a set of items in a serial learning task, or specific inter-item associations in a paired-associate task, unless he knows what items constitute the set. This means that the processes involved in free recall are in some ways more basic than those underlying many other laboratory tasks in memory, and that therefore the study of these processes constitutes a logical starting point for the study of memory. Secondly, free-recall phenomena were ignored by students of verbal learning and memory for a long time. As a consequence they have not yet been fitted into a relatively fixed theoretical framework as have been serial- and paired-associate learning phenomena. Thus, when we are faced with the problem of interpreting experimental findings in free recall, we are not hampered by extant theoretical notions. Thirdly, and most importantly, it is easier to find out in free-recall situations what the subject is doing when he is memorizing a set of items. The task set to the subject in serial- and paired-associate paradigms is too rigidly prescribed by the experimenter and all that can be observed at the behavioral level is how well the subject can meet the experimenter's requirements. In free recall, the subject is told to recall as many items as possible, but he is left free to recall the items in any order. Examination of the order in which he recalls the items provides considerable insight into the subject's covert activities in the process of memorization.

In multi-trial free-recall tasks, the same set of items is presented on a number of successive trials, the order of presented items usually varying from trial to trial (cf. Waugh, 1961), and after each input trial the subject recalls all items he can. When the number of items the subject recalls are plotted against trials, a typical negatively accelerated learning curve results (Miller and McGill, 1952; Murdock, 1960; Tulving, 1962a, 1964; Waugh, 1961). One of the questions I wish to raise is, why does such learning, i. e., improvement in trial-to-trial recall, occur. I have argued elsewhere (Tulving, 1962a, 1964) that learning in multi-trial free recall is a consequence of the fact that the subject organizes list-items into larger units of material, into subjective units or S-units. At any given time the subject can retrieve from memory only a limited number of units of a particular set of materials, but the larger these units, that is, the more individual items each unit contains, the larger the number of items recalled (Miller, 1956a, 1956b).

ORGANIZATION AND LEARNING

The original evidence for this contention was provided by the observations that the amount of organization subjects imposed on the material increased systematically over trials, accompanying the increase in the number of items recalled, and that when the amount of practice was held constant, there was a sizable correlation between recall scores and organization scores as subject variables. Such correlational evidence, of course, is not very compelling. One can justifiably argue that organization and recall are two parallel but otherwise unrelated manifestations of some common underlying mechanism. Some

writers (e.g., Asch and Ebenholtz, 1962) have indeed claimed that free-recall learning occurs independently of any inter-item associations, being a consequence of greater availability of items resulting from repeated presentation of items. To the extent that the development of S-units reflects the development of specific inter-item associations, the denial of the primary role of inter-item associations in free recall learning also constitutes a denial of the dependence of learning on organizational processes. How can we evaluate the hypothesized relation between organization and free-recall learning more directly than is possible on the basis of observed correlations between the two measures?

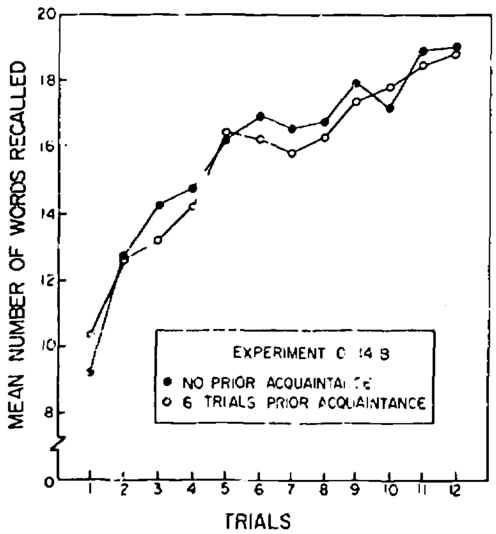
Two possibilities suggest themselves. The first is to present a set of items to the subject under conditions where he makes repeated responses to the items, but does not actively and deliberately try to organize the material into larger S-units. If learning depends upon organization, then such non-directed repetition should have little effect on recallability of items. The second possibility is to instruct subjects to use effective methods of organization. If learning depends on organization, such instructions would be expected to facilitate learning. We have done both types of experiments.

REPETITION AND ORGANIZATION

At first glance, it may seem difficult to create conditions under which subjects are repeatedly exposed to a set of items and when they make overt responses to these items, but do not think about and thus organize the items. Since we initially knew nothing about the problems involved, we decided to start with the most obvious approach, although I must admit, at the time we did not have much faith in it. Subjects were simply shown a set of common words, asked to read these words aloud a number of times, and then, immediately after such reading, were given the same set of words to learn under the typical free-recall instructions (Tulving, 1966). A control group learned the same set of words under identical conditions, but without having seen the words beforehand.

The list to be learned contained twenty-two high-frequency English words. In the "reading task," the subjects in one group were shown each of the words six times, in an uninterrupted sequence of 132 exposures. The words were exposed one at a time by means of a memory drum, at the rate of one word/second, and the subjects were instructed to call out each word as it appeared. The subjects in the other group went through the same procedure with a list of twenty-two names. Immediately following the reading task, all subjects were asked to learn the same set of twenty-two words to which the first group had been exposed already. Typical free-recall instructions were given, words were presented at the rate of one word/second, and at the end of each input phase subjects had 60 seconds for oral recall. Every subject received twelve learning trials. Thus, both groups learned the identical list of words under identical conditions. The only difference was that the subjects in the Prior Acquaintance Group has just finished reading aloud the same

FIGURE 1. Learning curves for two groups of subjects. One group received six trials of prior acquaintance with the set of twenty-two words to be learned later, the other group had no such prior acquaintance.



words on six continuous reading "trials," while the subjects in the No Prior Acquaintance Group had been "warmed up" with a different set of words.

The two learning curves from this experiment are shown in Figure 1. The Prior Acquaintance Group had a small, but statistically non-significant edge over the No Prior Acquaintance Group on the very first trial, but none thereafter. Over all twelve trials, the mean number of words recalled by the former group was 15.71, while the mean number of words recalled by the latter group was 15.91.

These results thus suggest that mere repetition of list items had no effect on subsequent memorization of these items. According to the present interpretation, *repetition had no effect because the subjects did not attempt to organize items in any way.* Post-experimental interviews revealed indeed that very few subjects had thought that they might be asked to recall the words they were reading in the original reading task. Most of them said they thought the experimenter was interested in how well they could read, how fast they could read or how clearly or consistently they could pronounce the words.

Although we did not give an immediate recall test to the subjects at the end of the preliminary reading task, it is extremely unlikely that they would not have remembered any words that they had seen. Thus, this experiment does not provide any evidence against incidental learning measured in a single recall test following a reading task of the kind we used. It only demonstrates that directed memorization under the conditions of free recall is not facilitated by whatever prior incidental learning has occurred. This is not a surprising result, if we assume that

no organization occurs during the reading task, that incidental memory of items read, therefore, would involve individual items as S-units, and that the subject can only recall a constant limited number of S-units on any given trial in free recall. If the subject, on the first trial in the directed learning task, recalls some of the items he learned incidentally, he could not recall too many other items because his limited retrieval capacity would already be partially exhausted. One of the implications of these findings is that if the subject, before the presentation of a longer list for directed recall, were told what some of the items in the list are, his recall would be no greater than that of the subject who has no such prior information, provided the items are not meaningfully related to one another and that the rate of presentation is rather fast. The hypotheses that increased recall can occur only as a consequence of organization and that the retrieval capacity is limited to a fixed number of S-units would completely account for this finding, even though it could not be readily predicted from many other theoretical points of view.

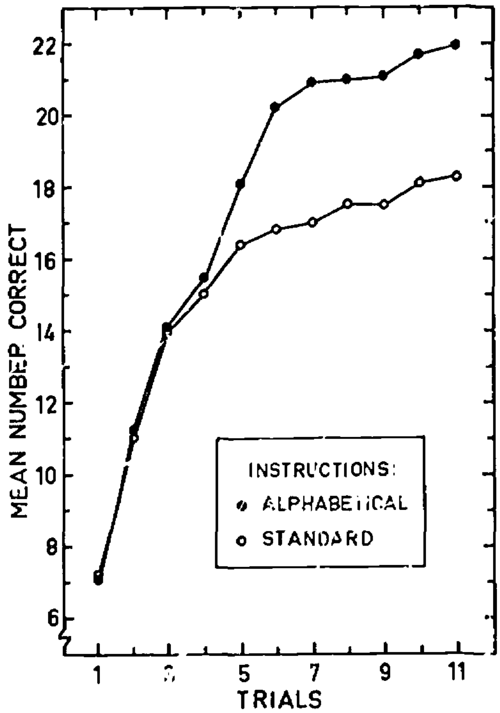
ALPHABETICAL ORGANIZATION

The second type of more direct test of the relation between learning and organization, as I said, involves manipulation of organizing strategies through instructions to subjects. An initial study of this kind was done as a classroom experiment (Tulving, 1962b) The subjects were eighty university students. They had to learn a list of twenty-two nouns, each beginning with a different letter, and the whole set covering a wide range of Thorndike-Lorge (1944) frequencies. In the input phase of each trial, words were projected on a screen at the rate of 1.5 second/word. In each output phase of 90 seconds duration subjects wrote down as many words as they remembered.

At the beginning of the experiment, all subjects were given typical free-recall instructions. For the first three trials all subjects worked under these identical instructions. Before the fourth trial, subjects were asked to open and read additional instructions on folded sheets handed out to them before the experiment. For half the subjects, the instructions simply reiterated what all subjects had been told at the beginning of the experiment, namely to "do your very best on each recall trial and put down as many words from the list as you can." The instructions for the other half of the subjects read: "Try to organize your recalled words alphabetically. When you look at the words on the screen, note their first letters, and make an attempt to associate the word with the letter. When you write the words down, go through the letters of the alphabet one at a time and try to remember the word that goes with each letter." Subjects in each group, the Standard Group and the Alphabetical Group, were given 30 seconds to read these additional instructions. The experiment was then continued under otherwise identical conditions for eight more trials.

The learning curves for the two groups are shown in Figure 2. For the first three trials, when both groups worked under identical standard instructions, the curves are almost indistinguishable indicating no differences in learning "ability" of the two groups.

FIGURE 2. Learning curves for two groups of subjects. After Trial 3, one group was given instructions to recall the words alphabetically, while the other group continued learning under standard instructions.



From Trial 4 on, however, the curves draw rapidly apart, the mean number of words recalled by the Alphabetical Group on Trial 5 equaling that of the Standard Group on the last trial of the experiment, Trial 11. Judging by the shape of the curve of the Standard Group, it appears that it would have taken them many additional trials to reach the level of performance comparable to that reached by the Alphabetical Group on Trial 6. The differences between the two groups would undoubtedly be even greater if it were not for the artificial limit on the performance of the Alphabetical subjects imposed by the "ceiling effect."

These results clearly indicate that methods of organization are very important in determining the course of memorization of a list of items. Alphabetical organization is apparently much more effective in this type of task than are other methods of organization – grouping of words by their meaning – used by the subjects in the Standard Group. The learning curve of the subjects in the Standard Group reflects their limited performance, but the limit is not wholly determined by their inadequate mnemonic ability. To a large extent, it is determined by their ignorance as to how to make best use of their mnemonic ability in this type of task. When subjects are given more powerful organizing instructions, their "ability" increases considerably.

One important implication of this observation – I am now going to digress somewhat from my main theme – has to do with the problem of individual differences in learning ability. It is commonplace to study individual differences in learning ability by dividing subjects into "fast" and "slow" learners on the basis of their performance in a learning task, and then look for similar differences between the two groups in various other situations, such as various psychometric tests. In the experiment I have just described, there were pronounced differences in the memorization performance between the subjects in two groups, but since the groups were selected randomly it is a foregone conclusion that they would have shown no differences on any other tests. How frequently are individual differences in learning "ability" determined by the same factors as those manipulated in the present experiment? The long history of failures to find powerful psychometric correlates of learning performance, or sizeable correlations between different learning tasks, suggests that individual differences in learning "ability" may be largely governed by individuals' knowledge of effective strategies of handling a particular task, rather than by the sheer goodness of their memory. At any rate, in the light of data from this experiment it seems that if we wish to add to our understanding of individual differences in learning "ability," a good place to start is to analyze the learners' strategies of memorizing in a given learning situation, and not to observe his behavior in some other extra-experimental situation.

ORGANIZATION AND LEARNING

Let me now return to the relation between organization and learning. We have seen that when a subject has to memorize a list of randomly selected words, alphabetical organization is more effective than other methods of organization. Alphabetical organization changes the task into a kind of paired associate learning task. The initial letter of each word serves as a cue for the recall of the whole word The experimenter does not have to present these cues since an articulate subject can generate them on his own. We could say that in alphabetical organization each input word is cued with respect to its initial letter and that its retrieval is facilitated by the availability of relevant cues. When other methods of organization are used, such as grouping of items in terms of their associative meaning, their belongingness to superordinate conceptual categories, or the degree of their prior familiarity, input words are cued with respect to their properties other than the initial letter and their retrieval is presumably effected through these other cues. It is possible, of course, that a given word is cued in several ways and that the way it is cued changes in the course of the experiment, but under the usual time pressure in a typical experiment, multiple cueing is probably an exception rather than a rule.

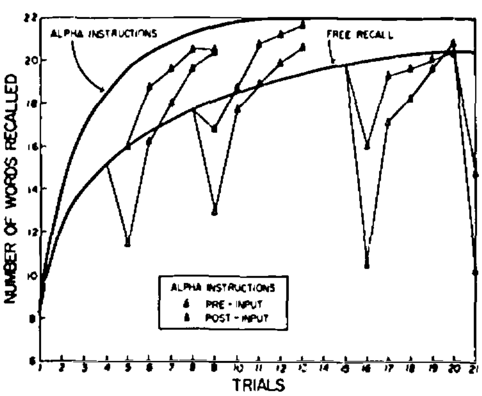
Suppose now that a subject has been exposed to a set of items for a number of trials, that he has not discovered alphabetical method of organization, and that he has been cueing the words in terms of their other characteristics. What would happen, if the subject were now told to switch to another method of organization, say alphabetical organization? We know already that alphabetical organization is likely to be more effective than other methods of organization and that therefore alerting the subject to it should facilitate learning. But it is also possible that changing the method of organization requires not only the adoption of the new strategy, but also the suppression of the previously used strategy. One might expect, therefore, that immediately after the introduction of alphabetical instructions the efficiency of the storing and retrieval mechanisms is temporarily reduced and that the disturbance is the greater the more highly developed the non-alphabetical organization.

Data relevant to these hypotheses come from an experiment done by Marcia Ozier as part of her PhD thesis (Ozier, 1965). Her subjects learned a list of twenty-two words, each beginning with a different letter. Words were projected on a screen at the rate of 1.5 second/word and subjects recorded their recall on paper-pullers that prevented them from seeing more than one word of their recall at any time. The duration of output phase of each trial was 90 seconds.

Most subjects were initially given standard free-recall instructions and then switched over to alphabetical instructions at some point in the experimental sequence. The two independent variables in the experiment were, first, the ordinal number of the "critical" trial on which alphabetical instructions were given, and second, pre-input versus post-input alphabetical instructions. Different groups of subjects were given alphabetical instructions on Trials 1, 5, 9, 16, and 21. In each of these groups, half the subjects were given alphabetical instructions before the input phase of the critical trial, while the other half received these instructions after the input phase, that is, before the output phase, of the critical trial. Before the critical trial, subjects in a given group worked under standard free-recall instructions. Thus, for instance, one group of subjects worked under free-recall instructions for eight trials, then received alphabetical instructions, followed by the input and output phases of the ninth and subsequent trials. Another group worked under standard free-recall instructions for eight trials, was then shown the words in the input phase of the ninth trial, and then was asked to start recalling words in the alphabetical order, beginning with the immediately following output phase. There was thus a total of ten independent groups, with twelve subjects per group. Scoring was lenient; even if the subject had deviated somewhat from the alphabetical order when he was supposed to recall the words alphabetically, he was given credit for each word correctly recalled. Subjects who obviously did not follow alphabetical instructions, however, were replaced.

The somewhat schematized results of this experiment are shown in Figure 3. The upper heavy smoothed curve shows the learning curve for the subjects who received alphabetical instructions before Trial 1, the lower heavy smoothed curve represents the free-recall conditions. Subjects who had received pre-input alphabetical instructions on Trial 1 had a mean recall of approximately eight words on that trial, while those who had received

FIGURE 3. The effect of alphabetical organizing instructions given at various points in the course of learning. See text for explanation.



post-input alphabetical instructions on Trial 1 recalled approximately five words, a difference of three words. The learning curve for the latter group is not shown on the graph. It reached the learning curve of the former group by the fifth trial, being practically identical with it thereafter.

The other data in Figure 3 show the initial drop and subsequent rapid recovery in recall of groups who were given alphabetical instructions at Trials 5, 9, 16, and 21. For each of these groups, with the exception of those for whom Trial 21 was the critical trial, mean recall scores are shown for five trials after the critical trial. These data show, first, that immediately after the introduction of alphabetical instructions there is a drop in performance – with the exception of pre-input instructions on Trial 5 – with the magnitude of the drop varying directly with the ordinal number of the critical trial and being consistently greater for post-input alphabetical instructions, and second, that the subsequent recovery of the effectiveness of alphabetical instructions given on the critical trial varies inversely with the ordinal number of the critical trial.

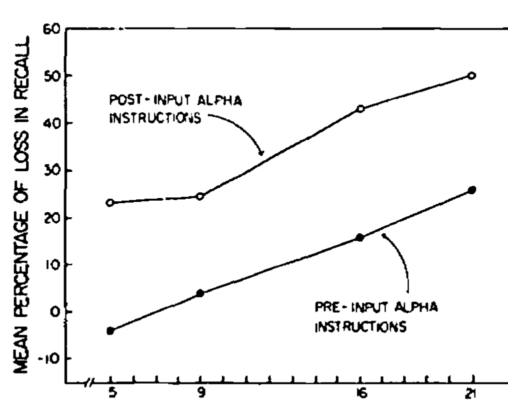
The percentage loss in recall in the first output phase following the introduction of alphabetical instructions is plotted as a function of the ordinal number of the critical trial, for both pre-input and post-input instructions, in Figure 4. Each data point in this graph represents the mean difference in recall between the number of words recalled on the critical trial and the immediately preceding trial divided by the mean number of words recalled on the immediate pre-critical trial. The data plotted in Figure 3 showed that the drop in performance after the introduction of alphabetical instructions, measured in terms of absolute number of items, varied directly with the ordinal number of the critical trial. The data shown in Figure 4 indicate that the same relation holds between percentage loss in recall and the ordinal number of the critical trial.

From previous experiments we already know that organization of material increases over trials of practice in the free-recall situation (Bousfield, Puff, and Cowan, 1964; Ehrlich, 1965; Tulving, 1962a, 1964). Therefore, manipulating the temporal point at which subjects are switched from other forms of organization to alphabetical organization essentially amounts to manipulation of the degree of existing organization that has to be changed. The results of Ozier's experiment suggest that the stronger the pre-existing organization, that is, the more firmly the list-items are cued with respect to characteristics of words other than their initial letters, the more disruptive is the requirement that the subject change his system of cueing and the longer it takes for the otherwise powerful alphabetical system of organization to become effective.

AVAILABILITY AND ACCESSIBILITY OF STORED INFORMATION

I have so far talked about two general classes of methods of organization: organization of words into larger groups in terms of their perceived, discovered, or fabricated meaningful relations to one another; and organization in terms of structural features of words, such as initial letters. In either case it is as if words stored in memory – or information about words to be reproduced – are in storage in a way that would facilitate retrieval: two or more related words are stored in or near the same location, such that when the retrieval mechanism finds one, the other one is also easily located, or else a word is stored in a location in which its initial letter is stored. Thus, one part of the stored information can be thought of as a retrieval cue for another part. Organization of information in the storage, however, can facilitate retrieval only to the extent that retrieval cues are available at the time of recall. Alphabetical organization is more effective than organization of items in terms of their semantic

FIGURE 4. Percentage loss of recall on the trial immediately following the introduction of alphabetical organizing instructions, as a function of the "critical" trial on which alphabetical instructions were given.



features because the initial letters of the words, serving as retrieval cues are readily available to the subject at the time of recall.

This kind of speculation implies that organization of material in the memory storage, or organized retention, does not necessarily increase the storage capacity, but rather makes the stored information more accessible to recall by providing distinctive retrieval cues through which individual items of information can be reached. It also implies that the amount of information potentially available in the storage at a given time may be greater than is apparent in typical unaided recall. In other words, limitation of memory is primarily determined by the limited capacity of the retrieval mechanism and not necessarily by the limited capacity of the storage mechanism.

In experiments of the kind that we have examined so far it is impossible to distinguish between retention and recall. If a subject using alphabetical organization can recall more words than an otherwise comparable subject who uses less efficient methods, we do not know whether this difference is attributable to differences in retention or only in recall. It is possible that both subjects have retained the same number of words and that the alphabetical organizer simply has access to a larger number of retained words, but it is also possible that the alphabetical organizer in fact has been able to store more information in his memory, that therefore more information is available in the storage, and that his superior recall merely reflects the greater amount of information that he has stored.

The logical requirements for the analytical separation of retention and recall, or the distinction between availability and accessibility of stored material, are simple enough. All that we need to do is to present the material to be remembered to different subjects under identical conditions, including identical instructions, and then vary experimentally the number of retrieval cues at the time of recall. If the groups are treated identically up to the beginning of the recall test, we can assume that the amount of information they have stored and thus the amount of information available is the same for all groups. If the number and kinds of retrieval cues given to different groups at the time of recall produce differences in recall, these differences would reflect differences in accessibility of the stored information and not differences in availability.

In the initial experiment we did to distinguish between availability and accessibility (Tulving and Pearlstone, 1966), we used a dual-level input list, consisting of category names and appropriate words as instances of these categories. Thus, for instance, one of the lists consisted of twenty-four words in twelve categories, with two words per category. The list presented to subjects might contain items like this: <u>things found on a farm</u> – WHEAT, TRACTOR; <u>substances for flavoring food</u> – SUGAR, CLOVES; <u>geological formations</u> – PLAIN, LAKE; <u>musical instruments</u> – DRUM, FLUTE; etc. Subjects were instructed that their task was to remember as many words as possible, that they would be tested for the words only, and that they did not have to memorize the category names. All subjects learning a particular list were treated identically up to the beginning of the recall test. At that point, half the subjects were asked to recall all the words they remembered, i.e, words other than the category names, while the other half were presented with the category names as retrieval cues and asked to write down all the words belonging to each category that they remembered from the list.

The design of the experiment included lists of three different lengths (twelve, twenty-four, and forty-eight words) and three levels of numbers of words per category (one, two, and four). Within each of these nine conditions of presentation, there were then two conditions of recall, cued and non-cued, as I have just described. The experiment was done with high-school students of both sexes as subjects, and data were collected in regular class sessions. Input material (category names and words to be remembered) was presented by means of tape recorder. The subjects wrote their recall in appropriately designed booklets. The presentation time varied with the length of the list and with the number of categories according to the formula: T= NoC+L, where T is presentation time in seconds, NoC is the number of categories, and L is list length. The amount of time given for recall was 1, 2, and 4 minutes for lists of twelve, twenty-four, and forty-eight words, respectively. A total of 949 subjects was tested, with approximately fifty-three subjects for each of the eighteen conditions of the experiment.

The mean number of words correctly recalled under each of the conditions of the experiment is shown in Table 1, together with corresponding standard deviations. As these data show, cued recall was superior to non-cued recall for all nine lists, the superiority varying directly with list length and inversely with the number of words per category. The difference between cued and non-cued recall thus was smallest for the list in which twelve words were organized into three categories, with four words per category. In this case, presentation of category names as retrieval cues apparently does not help subjects very much since they presumably can remember the category names on their own. The largest difference was found for the 48-word list in which each word belonged to a different category. Presentation of category names as retrieval cues in this list facilitated recall considerably: subjects in the cued recall group recalled over thirty-five words on the average, while the subjects in the non-cued group recalled less than sixteen words.

The major implication of the results of this experiment is that the subject's failure to recall a word he was instructed to recall does not necessarily mean that information about the membership of the word in the list is no more available in the memory storage. It is available, but not accessible under the typical unaided recall conditions. If the subject receives cues about the "location" of the relevant information in the storage, he will "find" many otherwise inaccessible words. Thus, intratrial "forgetting" (Tulving, 1964) of at least some of the items does not reflect the failure of the storage mechanism, but only the failure of the retrieval mechanism. How does the retrieval mechanism operate?

TABLE 1

MEAN NUMBER OF WORDS RECALLED AND THEIR STANDARD DEVIATIONS FOR THE EIGHTEEN EXPERIMENTAL CONDITIONS

List Length	Recall Conditions	Number of Words Per Category: 1 M	1 S. D.	2 M	2 S. D.	3 M	3 S. D.
12	Cued	10.70	1.57	10.94	1.12	9.98	1.55
	Non-cued	7.70	1.61	8.13	2.03	9.31	1.91
24	Cued	21.70	1.97	19.31	2.38	15.11	3.17
	Non-cued	11.18	2.56	11.82	2.54	13.38	3.08
48	Cued	35.35	6.42	35.76	7.86	29.60	6.00
	Non-cued	15.57	3.53	18.79	5.99	19.33	5.34

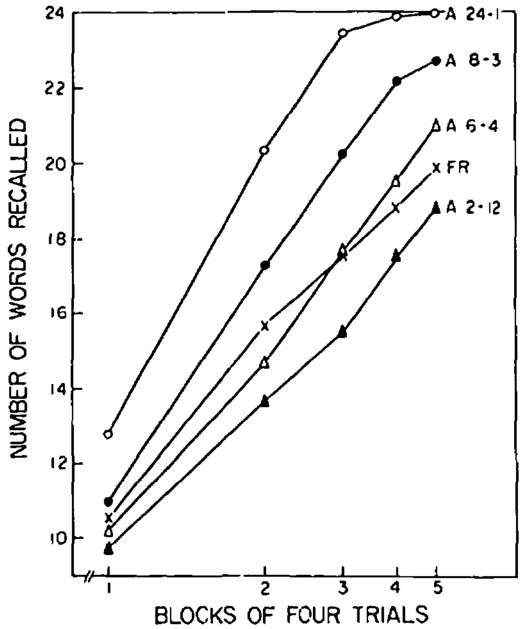
The Tulving and Pearlstone (1966) experiment provides some data relevant to this question. Word-recall data in an experiment of this type, where words are organized into higher-order units by the experimenter -- in this case, into superordinate categories -- and where this organization is adopted by the subject, can be analyzed into two independent multiplicative components, number of categories recalled and number of words recalled per recalled category (cf. Cohen, 1963, 1966). A category is said to be recalled -- to be accessible -- if at least one word from that category is recalled. Given the number of total words recalled and the number of categories recalled, the mean number of words recalled within recalled categories can be obtained simply by dividing the former number by the latter. For instance, if the subject recalls the words SUGAR, CLOVES, and LAKE, he has recalled two categories -- substances for flavoring food and geological formations -- and an average of 1.5 words per category, two words from one category, and one from the other.

The breakdown of the word-recall data into these two components revealed two interesting facts. First, presentation of category names as retrieval cues affected only the number of categories recalled. In all lists the mean number of words recalled within a recalled category was the same under the cued and non-cued recall conditions. Second, word-recall increased as a direct function of list length for both cued and non-cued recall conditions, but again this increase was attributable primarily to the increase in the number of recalled categories and not to the increase in the number of words recalled per category. For the 24-word and 48-word lists the number of words recalled per category was invariant.

These data suggest then that recall of words from accessible categories is independent of the number of categories recalled, regardless of whether differences in category recall are brought about through presentation of category names as retrieval cues or through increasing the list length and thus the total number of categories presented. Retrieval mechanism seems to operate in some sense independently at two levels: at the level of higher-order units of organization and at the level of individual items within these higher-order units. Limitation of recall is determined by limitation of retrieval at both levels.

A further demonstration of the relation between number of categories and recall is provided by another experiment of Ozier's (1965). In this experiment subjects learned lists of twenty-four words either under conditions of free recall or under conditions of alphabetical organization. For different groups of subjects the number of unique initial letters of words in the list was varied. For one group, all words began with different letters. For another group there were twelve pairs of words (words within each pair beginning with the same letter); for a third group, eight initial letters were used, each associated with three different words, and so on, to the group in which twelve words began with one letter and twelve words with another. Thus, the experiment can be thought of as an analogy to the Tulving and Pearlstone experiment, with list length held constant at twenty-four words, the number of categories (initial letters of words) varying over the range of two to twenty-four, and the number of words per category (words sharing the initial letter) varying correspondingly over a range of twelve to one. Ozier's experiment differed from the Tulving and Pearlstone experiment in that it was a multi-trial learning experiment. Each subject was given twenty trials of practice on the list to which he had been assigned. The order of words was randomized on all trials, the words were projected on a screen at the rate of one word/second, and subjects had 95 seconds for written recall on each trial. Under the

FIGURE 5. Mean recall on successive blocks of four trials for lists varying in number of different initial letters and learned under alphabetical organizing instructions (A), and for a list learned under free-recall instructions (FR). In the designation of A lists, the first figure refers to the number of different initial letters of words, and the second represents the number of words beginning with a common letter.



conditions of alphabetical recall, initial letters of the words were given as retrieval cues in the output phase. Subjects were premedical students, mostly men, with twelve subjects per group.

On the first trial, there were no systematic differences in the number of words recalled by different groups, and only after the first three trials did differences become noticeable. This suggests that the presentation rate of 1 second/word was too fast for words to be cued with respect to their initial letters. The learning curves for four alphabetical groups and a free-recall group are shown in Figure 5. The mean number of words recalled is plotted in Figure 5 against five blocks of four trials each on a logarithmic scale. These curves show that learning varied directly with the number of categories (number of unique initial letters) or inversely with the number of words per category. It is also interesting to note that free recall is more efficient than alphabetical recall of the list in which there are only two categories and twelve words per category.

The results of the last two experiments suggest that under conditions where the subject is provided with efficient retrieval cues - or where he can easily generate them from his own memory - and where the inter-item associative strength of items within a category is minimal, the optimum strategy is to organize the items into a maximum number of categories. On the other hand, under conditions where the subject can benefit from organization of individual items into higher-order units only to the extent that he himself remembers what these higher-order units are, the optimum strategy is to organize items into relatively few categories. It is most likely, however, that when the number of categories is decreased beyond a critical limit the difficulty of retrieval of items from within a given category will outweigh the facilitating effects of smaller number of categories, and recall will decrease. In the limiting case where all items in a list belong to a single category, recall is no better than recall of a list in which each item belongs to a different category.

SUMMARY

I have argued that understanding of many phenomena of verbal learning and memory would be enhanced if we adopted the basic orienting attitude that a person's mnemonic performance depends to a very large extent on the methods that he wittingly or unwittingly uses to organize the material. The important function of such organization is to facilitate retrieval of information from the memory storage.

I considered in some detail the problem of the effect of repetition on recall of a set of discrete items. Mere mechanical repetition of items does not seem to have any noticeable effect on the subsequent directed memorization of these items, provided that individual items as such are already well known to the subject before the experiment. Learning of the set, however, can be considerably speeded up if the subject uses effective methods of organizing the material. Individual differences among subjects in various learning tasks may reflect primarily their knowledge of, and the ability to utilize effective strategies and mediational devices, and only secondarily their innate differences in memorizing ability.

Different methods of organizing material may be incompatible with one another. When the subject has been using one method and then is asked to apply some other method to the material he is memorizing, the adjustment to the new method requires some time. Interference with overt recall under these conditions reflects incompatibility of different types of organization, and adds further support to the hypothesis that learning is critically dependent upon how the material to be remembered is organized.

It is not necessary to assume that organization changes the amount of material that the subject can retain, or the amount of information potentially available in the memory storage. Evidence suggests that organization only serves to make more of the available information accessible. The learner has always more mnemonic information available, relevant to the successful accomplishment of a memory task, than he can ordinarily make accessible to recall. Accessibility of information depends not only on its availability in the storage, but also on retrieval cues. Retrieval cues may be part of the environmental conditions prevailing at the time of recall, or they may have to be carried in memory. Limitation of the retrieval mechanism

thus seems to be determined by the lack of accessibility of efficient retrieval cues, and the functional significance of organization lies in its role of providing such retrieval cues.

Optimum organization depends on the requirements of the recall task. If retrieval cues have to be generated from memory, the material is best organized into a relatively small number of categories - Mandler (1967) suggests that this number for common words is approximately five - if, on the other hand they will be available in the perceptual environment at the time of recall, the number of categories should be as large as possible. Retrieval cues such as category names provide access to categories or other kinds of higher-order units of material, but once access to such higher-order units becomes possible, retrieval of individual elements within units is unaffected by the number of higher-order units already recalled or yet to be recalled.

The evidence on which this account has been based is admittedly rather fragmentary, but the initial results are quite encouraging. Certain consistencies have already become apparent and further work will probably clarify those aspects of the conceptual framework that are still obscure.

While all this research that I have discussed was not undertaken with a view to solving practical problems as they occur in learning and remembering in real-life educational settings, it does have some implications for these problems. I will review just one, fully realizing that it is simply a reiteration of what many people have said before.

The kinds of findings from the simple experiments I have described in this paper once more underscore the need for the examination of methods that learners in classrooms use when they acquire knowledge and skills. We can never be certain that the methods of learning a person uses are the optimum ones for the requirements of any given task, or that these methods are sufficiently flexible to fit all tasks which in fact do require different strategies. And at the present time, we probably could not even evaluate the appropriateness of different approaches to different tasks or different subject matters, since we know so little about which methods are optimal for which kinds of learners in which situations. It seems to me that educators and psychologists have been more interested in methods of teaching than in methods of learning. Teachers are frequently taught how to teach, but how often are learners given specialized instruction in how to learn? A good teacher may be successful because, among other things, he organizes the material to be learned for the learner, but a good learner is likely to be more successful because he can organize the material on his own even if it is not optimally structured by the external environment.

To the extent that methods of teaching and methods of learning can be separated, it would seem that in the final analysis the latter are more important than the former, if for no other reason then at least because it might have greater transfer value as far as the individual learner is concerned. Even a poor learner, one who does not know how to use his innate intellectual abilities most effectively in learning situations, can do better if he has a good teacher to organize the material for him, but he cannot always expect to have good teachers around him. In many situations he has to be his own teacher. Would it be too rash to suggest that instruction in how to learn and how to remember be as important a part in the curriculum of the schools of tomorrow as are the three R's now?

FOOTNOTE

1. Research reported in this paper has been supported by Grant No. APT-39 from the National Research Council of Canada, and Grant No. GB-3710 from the National Science Foundation.

REFERENCES

Asch, S. E., and Ebenholtz, S. M. The process of free recall: Evidence for non-associative factors in acquisition and retention. Journal of Psychology, 1962, 54, 3-31.

Bousfield, W. A., Puff, C. R., and Cowan, T. M. The development of constancies in sequential organization during repeated free recall. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 489-495.

Bugelski, B. R. Presentation time, total time, and mediation in paired-associate learning. Journal of Experimental Psychology, 1962, 63, 409-412.

Cohen, B. H. An investigation of recoding in free recall. Journal of Experimental Psychology, 1963, 65, 368-376.

Cohen, B. H. Some-or-none characteristics of coding. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 182-187.

Ebbinghaus, H. Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Leipzig, Drucker und Humblot, 1885.

Ehrlich, Stéphane. Le rôle de la structuration dans l'apprentissage verbal. Psychologie Française, 1965, 10, 119-146.

James, W. The principles of psychology. Vol. 1. New York: Holt, 1890; reprinted by Dover Publications, Inc., 1950.

Mandler, G. Subjects do think: A reply to Jung's comments. Psychological Review, 1965, 72, 323-326.

Mandler, G. Organization and memory. In K. W. Spence and Janet T. Spence (Eds.) The psychology of learning and motivation. New York: Academic Press, 1967.

Miller, G. A. Human memory and the storage of information. IRE Transactions on Information Theory, 1956, 2, 129-127. (a)

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97. (b)

Miller, G. A., Galanter, E., and Pribram, K. H. Plans and the structure of behavior. New York: Holt, 1960.

Miller, G. A., and McGill, W. J. A statistical description of verbal learning. Psychometrika, 1952, 17, 369-396.

Murdock, B. B., Jr. The immediate retention of unrelated words. Journal of Experimental Psychology, 1960, 60, 222-234.

Ozier, M. Alphabetic organization in memory. Unpublished doctoral dissertation, University of Toronto, 1965.

Pollack, I., Johnson, L. B., and Knaff, P. R. Running memory span. Journal of Experimental Psychology, 1959, 57, 137-146.

Thorndike, E. L., and Lorge, I. The teacher's word book of 30,000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Tulving, E. Subjective organization in free recall of "unrelated" words. Psychological Review, 1962, 69, 344-354. (a)

Tulving, E. The effect of alphabetical subjective organization on memorizing unrelated words. Canadian Journal of Psychology, 1962, 16, 185-191. (b)

Tulving, E. Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. Psychological Review, 1964, 71, 219-237.

Tulving, E. Subjective organization and effects of repetition in multi-trial free-recall learning. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 193-197

Tulving, E., and Pearlstone, Zena. Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 381-391.

Underwood, B. J., and Schulz, R. W. Meaningfulness and verbal learning. New York: Lippincott, 1960.

Waugh, N. C. Free versus serial recall. Journal of Experimental Psychology, 1961, 62, 496-502.

Developmental Processes in Thought

Kenneth Lovell
University of Leeds
Leeds, England

WHEN Professor Klausmeier chose the title for my paper, he asked me to deal with three points. First, give an outline of the conceptual framework that I am going to use to indicate developmental processes in thought. Second, give some account of the research that has been carried out to test the validity of the general framework. Third, point out the educational implications of what is dealt with in the first part of the paper.

The treatment of intellectual processes, including thought, is rather different in American psychology from that in either Russian or Swiss psychology. Only the future can show the extent to which these treatments can enrich one another, or the extent to which the different approaches say the same thing using different terminologies. Individuals wishing to know more of these three approaches would, no doubt, be interested in Berlyne's most recent book Structure and Direction in Thinking (1965). However, I shall deal with my topic largely from the standpoint of the Genevan school, for Piaget has great influence in British educational circles, although he is not quite so influential among British psychologists.

FIRST 21 MONTHS OF CHILD'S LIFE

There is now much data which indicates that the first 21 months or so of a child's life are of great importance from the point of view of the growth of thought. Piaget has shown that a baby's movements are not random; indeed, on close inspection they are seen to conform to a pattern. This period of development was called by Piaget the period of sensori-motor intelligence, because the schemas, that is the sequence and the structure of the actions slowly built up in the mind, are dependent upon the direct support of perception and motor activity. From 12 to 18 months of age the child goes through a period of intense experimentation. He will let an object fall to the floor to see what happens; he will knock things, shake them, or throw them to see if they roll. New ways of doing things are discovered based on a grasp of new relationships. For example, a child will pull a box toward him in order to obtain a toy placed on it, which otherwise would be out of reach. And between 18 and 21 months of age, the awareness of relationships is sufficiently developed for the child to be able to invent new means, and to be able to foresee which acts will succeed and which will not, without putting them to the test. This invention comes through a covert process that amounts to internal experimentation. The earlier trial and error period is not now so important, for the child can increasingly represent to himself the various possible actions and how they must be combined, to attain a desired end. The child is beginning to think, for his actions are now carried out in an inward form. Some psychologists would call such internalized actions, incipient or implicit responses.

During the first 21 months or so of life, the ordinary child elaborates the basic schemas of the object, of space, of time, and of causality. For example, the basic schemas in relation to space and time are laid down as when the child adjusts his reaching actions for near and distant objects, and when he moves to catch a swinging rattle.

In Piagets's view, the child in adapting himself to his environment has to assimilate or absorb new experiences into his existing schemas. At the same time he must accommodate to the environment through the modification of existing schemas or the build up of new ones. Cognitive adaptation to some new situation, and hence the ability to deal with it in thought, involves both these processes. To assimilate the meaning of a new situation the child must accommodate to it; and in order to accommodate to it, he must be able to assimilate it. Cognitive growth is thus a step by step process with the new always building on the old. But once new experience is assimilated, the child's schemas become more complex, and because of this, more complex accommodations are possible. Moreover, the child's schemas do not remain unchanged even in the absence of environmental stimulation, for meanings are constantly reorganized, and linked with other meanings. This internal renovation is, in Piaget's view, an important source of cognitive development. Even so, intellectual growth is slow. We must also note that Piaget's observations lead him to believe that it is schemas in the process of organization that children tend to repeat playfully and with seeming pleasure. But when such schemas have become organized, the apparent pleasure disappears and they cease to be repeated unless they are combined to form new schemas or serve as a means to some end. Flavell (1963) remarks that schemas are structures, and one of their important, built in properties, is that of repeated assimilation of anything assimilable in the environment.

BETWEEN 2 AND 8 YEARS OF AGE

Throughout the sensori-motor period the child is unable to use an image or word, which could represent to him an object or event not actually present. For Piaget, the child moves out of this stage into what he calls the pre-operational stage of thought when the child can differentiate a word from what it stands for; for example, use the sound "dog" to stand for, or represent, a dog. When he can do this he can represent to himself situations that are no longer in actual evidence, and thought is lifted to an entirely new level. Note carefully, however, that thought comes before language; the latter is fitted on to thought that already exists. But once the child can use language, thought is extended over an immensely increased range. Moreover, he now has a more permanent and far more flexible model of the outside world, for he is no longer dependent upon immediate perception and motor activity for thought. Nevertheless, between 2 and 4 or 5 years of age, thinking tends to center on one striking feature of the situation; it is also irreversible in that the child is unable to move back, in his mind, to the starting point from which his immediate thinking began. The child at this stage can only see an event from his own viewpoint; he is unable to conceive the viewpoint of other children. Yet by 3 or 4 years of age the child has copied the adult model of language, and at this age we are in grave danger of overestimating his level of thought from the nature of his speech.

By about 5 years of age a change begins to set in, and by 7 to 8 years of age the child's thinking becomes more systematized; that is, his thoughts now conform to certain rules and his thinking becomes what we adults call logical. The sequence and structure of actions in the mind, or the schemas, now available are now altogether different in kind. The ability to reason and "understand" demands higher order schemas which permit a simultaneous grasp of the successive sequences of actions taking place in the mind. The ordinary child of 7 to 8 years can, as it were, "look down on his schemas" or "turn round on his schemas." He is then aware of the sequences of action in his mind; he can see the part played by himself in ordering his experiences; and for any action in his mind, he can see that there are other actions that give the same result. That is, he sees equivalences. Thus he understands that $4+2=3+3=7-1$ or that all girls plus all boys plus all adults = all boys plus all adults plus all girls. Indeed, there is now a new kind of coordination of schemas, yielding a simultaneous understanding of equivalences for actions within the mind. Thus the child can now measure the same distance in feet and inches and understand that the different figures mean the same thing; he can perform subtraction using the method of complementary addition; he can decompose 42 units in 4 tens and 2 units with understanding. Thinking now conforms to a system, and there is learning with a certain amount of understanding.

At the level of sensori-motor and pre-operational thought there was a great deal of learning, but little or no understanding in the sense that the child learned a linear sequence of actions but was unable to elaborate a set of equivalences. We must not in any way belittle learning that takes place before thought becomes systematic, for such learning is essential for the growth of thought itself. Indeed in older normal children, and in adults, a great deal of learning is of this type.

Now because the elementary-school child can increasingly see the part played by himself in ordering his experiences, and because he can increasingly coordinate actions in his mind and make them conform to a system, he can build the concepts of a class, a series, length, time, and so on. But the concepts that he builds are only those that he can attain from experience with first hand reality, and his systematized or logical thought is related only to the world of perceivable things and events. For Piaget, the elementary-school child is at the stage of concrete operational thought and is, for example, able to understand certain aspects of mathematics.

END OF ELEMENTARY SCHOOL PERIOD

Toward the end of the elementary-school period, however, children begin to realize that there are gaps and uncertainties in their thinking, and that certain kinds of problems cannot be solved by them. As the child becomes better at organizing and structuring problem data with concrete operational methods, he becomes more aware that the latter does not yield a logically exhaustive solution to his problem. He gropes for new methods of attack, often in Piaget's view, as the adolescent commits himself to real life situations.

So from 11 to 12 years of age in able children, and from 13 to 14 in ordinary pupils, new thinking

skills begin to emerge. Due to his continued interactions with the cultural milieu, and to the maturation of the central nervous system, the individual can produce more complex expectations when faced with certain kinds of situation and data. His logical thought is no longer restricted as before, for he can increasingly use statements or propositions relevant to objects and their relations. The sequence and structure of his mental actions, or his schemas, are new in kind. The pupil can now begin to manipulate statements which refer to classes and relations. For example, he can answer the question "If Henry is taller than Mary and shorter than Debbie, who is the tallest?" Or he can argue from the situation in front of him that, "Either this must be true or that must be true," or he can see that "X implies Y." The pupil can, for any sequence of actions in his mind, see a greater range of equivalent actions than was the case at elementary-school level. Faced with certain situations the child can, at 13 or 14, set up an hypothesis and work out what would happen if this hypothesis was true. He is now able to deal with the merely possible. The pupil has reached what Piaget calls the stage of formal operational thought, and many experiments to illustrate the growth of such thought are to be found in Inhelder and Piaget's (1958) book The Growth of Logical Thinking from Childhood to Adolescence (1958). A number of these experiments were repeated by Lovell (1961) and Jackson (1965), and the broad stages in the growth of logical thinking were mainly confirmed although the position is more complicated than Inhelder and Piaget suggested. It must be stressed, however, that formal thought is found in all areas of thinking - in history, literature, politics - and not just in science and technology. There is much published work to support this statement, and if space permitted, I could give you some interesting illustrations for formal thinking in these fields.

With the onset of formal thought, the pupil is able to elaborate an entirely new kind of concept. You will remember that in the elementary school the child could say how he ordered his experience, for he could dissociate the part played by himself in classifying his experience from the characteristics of that experience. But in adolescence the pupil is able to structure and coordinate actions upon relations, which themselves result from the coordination of actions. For example, in the case of heat in physics, the concept depends on the earlier elaboration of the concepts of mass and temperature. The latter are developed at the level of concrete operational thought, for each is a coordination of some intuitive aspect of reality, but their product is not. Concepts derived at the level of formal thought depend upon the concepts elaborated at the level of concrete thought being completely detached from their concrete contexts and manipulated as "pure" concepts.

BEHAVIOR THEORY

Many individuals have been brought up on what may be broadly called behavior theory, although not all Americans think of cognitive growth in behavioral terms. Some are now talking of "strategies"; "structure" and "rule"; "programs"; "plans." But for people who have studied behavior theory I would like to review a few points that will help them grasp how the simplest principles of behaviorism look to a Piagetian. First, if two schemas are frequently evoked in regular succession they tend to form a single, larger, schema comprising the first and then the second in that order. The new and larger schema is then evoked as a whole. This is "conditioning." The individual action is, of course, the limiting case of a schema. Second, if two or more schemas are frequently evoked in irregular succession they, too, tend to form a larger schema made up of all the original schemas. This larger schema will also be evoked as a whole, although the order of the evocation of its parts will depend upon the stimulus properties of the situation. Nevertheless, its value, as a larger schema, lies in the mutual inter-facilitation of its parts. This provides the basis for "trial and error" behavior thought by Thorndike to be the prototype of all learning. Third, if a situation evokes a larger "trial and error" system of actions or schema as indicated in my second point, and the larger schema includes a particular sub-schema that is particularly appropriate to the situation, the effect of accommodation will be to inhibit and finally eliminate the other sub-schemas and so strengthen the particular sub-schema (Thorndike's Law of Effect).

The functional principles indicated under my three points seem to be common to all schemas and show the most elementary forms of coordination of schemas. They may account for habit formation, or for simple kinds of learning, either in humans or in animals, which involve a linear sequence of actions. But we do not know if they could be made to account for understanding, for the capacity to reason and to understand are not functions or properties of all schemas. From time to time a combination of my first two points may well give more "intelligent" trial and error behavior from which outsiders or onlookers may infer that two separate sequences of actions are equivalent to one another. For example, a young child, or an older retarded child runs into the house by the shortest route which is the one normally used. But if he finds the gate locked, he at once selects the best alternative. Equivalence is implied by the detour behavior, but it is not explicit in the child's representation of the situation.

Within these first and second order operational schemas, i.e., schemas that function at the level of concrete and formal operational thought respectively - the operational structure of systems of equivalences form a sub-schema of great generality and enter into all processes of learning. The acquisition of such schemas are essential to learning. Such schemas do not arise merely by the pupil knowing if his answers are right or wrong; rather they seem to depend upon an active coordination by the subject of his existing schemas - on his ability to order these actions simultaneously and to recognize their systematic equivalences. Good teaching presents situations in which such coordination is possible and necessary, and in enough variety for the pupil to dissociate what is common, although it cannot ensure this. It will readily be realized that not all learning either in elementary or high school is of this form. When a 9- or 15-year-old is presented with a problem for which he has an adequate schema, he will assimilate the problem to the

schema by an appropriate method of attack and arrive at a solution. This is learning of content. He has not acquired a new method of coordination for he already had the method. All the problem did was to strengthen the tendency for this schema to be evoked when faced with problems of this type.

The most elementary ways in which schemas can be coordinated have just been indicated and all schemas probably lend themselves to this form of coordination. But since these simple methods of coordination do not involve a simultaneous apprehension of successive schemas by higher order schemas, the connections implied by the coordinations are linear. We thus have a linear series of actions at the first level of abstraction involving "learning" but little understanding. Most learning in the mentally retarded, and much in ordinary humans, is of this type.

Let me emphasize that I am well aware that I have not allowed myself to get bogged down over the question of whether or not the stages of cognitive growth that I have indicated do actually exist in the child's intellectual growth. Piaget has been much criticized over his stages, some having argued that stages do not exist in other aspects of growth. The current practice among researchers of subjectively deciding whether responses can be fitted into Piaget's schema of stages is obviously not suitable for assessing the validity of these stages. More objective methods are required which would allow the data to determine their own patterns and groupings and to confirm or deny the existence of stages.

FINDINGS FROM RESEARCH ON PIAGET'S IDEAS

The second part of my paper is about the findings from the research that has been carried out to test Piaget's ideas. I have heavily condensed the literature above since the material now at hand is very great. Studies have been made in the realm of the general growth of logical thought, into the child's ability to classify and serialize, and into his notions of space, time, causality, and the like. Whatever the future holds regarding the presence of stages, it can be said at the present that research shows that a high proportion of protocols obtained from pupils can be put into the stages proposed by Piaget, or into intermediate stages. It is true to say that the broad picture of the growth of thought, presented by Piaget, has been confirmed, although there are many blots on the canvas, and at times one must be extremely critical of his work. But all in all he has done more than anyone else to throw light on the growth of thought processes. His experiments will stand the test of time, although his theoretical system is certain to undergo changes.

The great number of studies relating to the growth of concepts shows that it is between 8 and 8 years of age that the pupil begins to elaborate the basic concepts of mathematics and science for he can now dissociate the part played by himself in ordering his experience. If we take, say, the concept of time, it has been amply confirmed that the child begins to coordinate instants and intervals around 8 years of age, although he may well have been using time words and telling the time for some years previously. Likewise experiments relating to the overall growth of logical thought have shown the same general progression whether in the area of science, history or religious thinking. The available data again confirms that it is not until 13 to 14 in ordinary pupils and 11 to 12 in the very able, that formal thought is possible and second order concepts are elaborated. I must stress, however, that the stages are not clear cut, and I shall return to this point. Indeed the stages are blurred, so that when a pupil is approaching the stage of concrete or formal thought, he is found to be at a more advanced stage of thinking in one situation than he is in another although both involve the same operational structure. It is also true that in some experiments certain stages appear to be absent; some responses are found that cannot be fitted into Piaget's stages; and occasionally a child is found who appears to conserve, say, weight before substance. On the other hand, there is no research that I know of, that has shown that Piaget was wrong concerning the sequence of stages for any one situation for any one child. This was clearly seen in a recent study, just completed, where a group of boys was tested on the same experiments each year between 11 and 15 years of age.

It would not be inappropriate if I now gave you a couple of examples of protocols obtained in a recent study in the growth of logical thought in relation to history. It will keep our discussion well and truly about children. I take the example from history as I cannot then be accused of being interested only in mathematics and science. Pupils read a story concerning William the Conqueror – the last man who successfully invaded England and that was in 1066. With the story still in front of him, each child was asked individually, "Was William the Conqueror a cruel man?" This was followed by "Why do you think so?"

One pupil age 13 years 8 months gave a reply that was clearly at the level of concrete thought. He said, "Well, William wasn't cruel at the beginning because he allowed the English to collect the dead bodies. If he was really a cruel man he wouldn't even have allowed that. Yet later on, after the Danes had left, he'd no need to take revenge. He could have reconquered the North – tried to make treaties and things." Here the pupil could use the information provided, but he was not able to form a mature hypothesis from a consideration of all the implications in a situation.

Another pupil age 14 years and 8 months replied at the level of formal operational thought. "It depends on what you call cruel. If the definition of 'cruel' is to kill and ravish and burn for any purpose whatever, William was cruel. On the other hand, if one is prepared to accept political necessity, William's cruelty was justified. Compared with many other feudatories, knights and so on, he was essentially a kind man. They ravaged generally for their own advantage and without care for the common folk of the land. Duke William, if the common people went with him, seems to have been prepared to protect the common people from ravages. If, however, they went against him, he seems to have treated it as a deliberate breaking of faith and acted accordingly. So, by the standards of his own day – for we really cannot judge him by our standards –

he was probably not a 'cruel man.'" This pupil was clearly able to reason by implication of an abstract lead.

One of the consistent findings in our studies at Leeds, and it has been supported by the work of Jackson (1965) - a student of Dr. Lunzer of Manchester - is that only a proportion of school educable retarded pupils reaches the level of concrete operational thought even at the age of 15. It has repeatedly been shown to be true in respect of mathematics and science; while recently we have shown that the growth of English morphology in such children, even at 15, is less advanced than that of normal children of 4 and 5 years of age if we accept Berko's (1958) data as representative of American children. Our findings, which will be published, were predicted from our knowledge of the thinking of this type of child in other areas. The learning problems of school educable retarded children, and of the least able left in elementary and high school, stem from the fact that their schemas permit only a linear series of actions below, or just at, the level of concrete operational thought. They learn, but they do not "understand." Transfer certainly takes place in these pupils as it did in Harlow's apes, but only when the schemas involved are less complex than those required at the level of concrete operational thought.

One great issue that research has clearly indicated is that problems that appear to have the same operational structure are not all solvable at the same time. Much research shows that both concrete and formal operational thought is, at first, a task specific within limits, and that thinking is not organized to the extent that Piaget's theory would indicate. Dodwell (1960, 1961) showed that there was only a moderate correlation between tests which, in Piaget's view, all involved the same level of thinking and which all involved integral aspects of the number concepts. Later Dodwell (1962) pointed out that while the concepts of class and number develop within the same age range, there was no clear indication that they both arose together or that one was elaborated before the other. Likewise we at Leeds have found, as did Dodwell (1963), that a pupil could be at different stages of thought in three tests dealing with the concept of axes of reference. Indeed, Piaget (1960) himself has now admitted that operations are only gradually applied to larger and larger numerical sets, while he has always made it clear that it takes a child, on the average, two years to generalize concrete operations involved in appreciating conservation of quantity before they can be applied to conservation of weight, although from the point of view of cognitive structure the operations are the same. Even when the concept involved remains precisely the same, the quality of the child's thinking will vary according to the apparatus used (cf. Lovell and Slater, 1960).

It would, of course, be unfair to the Genevan workers if one failed to emphasize that they have made some provision for these eventualities. Inhelder and Piaget (1958) pointed out that concrete operations consist of the direct organization of immediately given data and they cannot be generalized to all situations at once. For example, length is conceived before weight. This, in the view of Inhelder and Piaget, is because it is more difficult to serialize, equalize, etc., objects whose properties are less easy to dissociate from one's actions, e.g., weight, than to apply concrete operations to properties that can be rendered more objective, e.g., length. Piaget (1956) also speaks of the notion of "horizontal differentials." This suggests that the same or similar concepts when derived from different materials or situations, develop in staggered sequence rather than simultaneously. But this notion does not fit well into his general theory.

At the level of formal thought we find much the same. In a study recently completed, in which children were questioned individually, we have found that the schema of proportion is not available in problems involving money, speeds, areas, series, etc., at the same time. Mathematical concepts seem often to be available at first in specific situations, even when they depend on second order operational schemas. Lack of specific experience, information, vocabulary, expectancy or individual differences in intellectual functioning which are unknown, most probably all play some part.

A number of studies have been undertaken, based on a neo-behaviorist approach, to see if intensive periods of specific training lasting a few weeks, can speed up the growth of understanding of particular issues; for example, the conservation of number or weight. The outcome of these studies has been largely inconclusive. Piaget's view (Ripple and Rockcastle, 1964), which was expressed in 1964, is that the child may learn something of the situation, but the training will have no effect on his general level of understanding for the specific attack is too trivial. The modification of a child's mental structures in Paiget's view, necessitates a far wider, more lasting, and more radical approach which involves many of its child's activities.

THE WORK OF SMEDSLUND

In concluding this section of my paper I will review a few details of a recently published work of Smedslund (1964). It was a very important study as it is extrememly relevant to some of the points that I have raised in this section of my paper. The specific purpose of the study was to investigate the interrelations of the specific acquisition of ability for concrete reasoning using different items. He carefully laid down a number of methodological rules which were applied to the construction of items in order to maximize their diagnostic validity. The items included class inclusion, reversal of spatial order, conservation of discontinuous quantities, conservation of length, transitivity of length, etc., - nine items in all. Many of the items have sub-items. All the tests were given individually to 160 children age 4 years 3 months to 11 years 4 months, evenly distributed over age and sex.

The upshot of his findings was that under the given conditions of the experiment, one could predict with a fairly high degree of confidence from the success or failure on one sub-item to success or failure on another sub-item; that is, intra-item reliability was high. On the other hand, the picture was different between items, for the inter-item

relationships were far lower. The difficulty of the items varied from 122 passes to 77 passes. When children who had passed or failed all the items were excluded, only 18 percent of the remaining subjects conformed with the hypothesis that the items are acquired in the order of their difficulty as measured by the total number of subjects passed.

Smedslund points out, in retrospect, what he considers to be a marked weakness in his study. He proposes that when investigating a concrete reasoning task we need to make a clear distinction between percept – a process depending upon the momentary stimulus input; goal object – that which the subject is told to attain; and inference pattern which is formed by the set of premises and the conclusion. For example, the stimulus situation as apprehended by the subject are the percepts; quantity and length are goal objects; and conservation and transitivity are inference patterns. Smedslund thinks that since all three factors influence the solution rate of items, the effect of a single factor can only be studied with the others held constant. The author is of the opinion that other things being equal, conservation precedes transitivity, but he feels that only in relation to one goal object, and one narrow perceptual context, will it be possible to determine whether or not conservation, transitivity, seriation, associativity, commentativity, addition/subtraction, multiplication, etc., are acquired in rigid predictable sequences. Here, indeed, is a field for research.

IMPLICATIONS

In this section of my paper, I want to discuss my interpretation of the educational implications of Piaget's work and to suggest one or two lines of research. To be helpful I have listed a number of points.

1. While experience is always necessary for mental growth, Piaget is clear that mere submission to external experience is not sufficient to make a child restructure his own thinking. Rather the child has to be active. He has to act on material things and become aware of the significance of his actions.

2. His observations led him to believe that it is schemas that are in the process of organization that children tend to repeat playfully and with seeming pleasure. Further, when such schemas have become organized, the apparent pleasure disappears and the schemas cease to be repeated unless they are combined to form new schemas or serve as a means to some end. Learning then seems to start from the child, from the schemas that he already has available. Actions on his part, that is, exploring, discovering, using new ways to solve old problems, all have an intrinsic interest for the young, and are self-extending. Naturally the level of intrinsic motivation varies from child to child, due, perhaps, to the level of activity in certain areas of the central nervous system. But when the schemas required for the solution to some problem are not too far removed in complexity from those available to the child, the inadequacy of existing schemas will force him to accommodate to the conditions of the problem. Hence the child restructures his own schemas toward greater cognitive adaptation to his environment. Not only does the child solve the problem, but he extends his capacity for further learning. The implication of Piaget's findings is that we need to begin to stimulate the intellectual growth of the child in the early weeks of life. The child needs to encounter a variety of changes of sensory experience in respect of sight, sound, touch, and movement. He needs an opportunity to see and grasp a variety of objects, and later, suitable toys that involve him in a greater range of actions; opportunities for physical movement and experimentation; and novelty. We must always bear in mind that there must not be too great a gap between the schemas available to the child and those demanded by the situation. Yet in spite of the help given by Piaget in assessing a child's level of thinking, what is involved in producing the correct amount of gap between the schemas available to the child and those demanded by the situation remains vague. This is where the intuitive skill of the teacher is called for. It is his task to arrange, or find in the environment, problems which call forth the schemas of the child in new and novel ways.

3. While language acts as a vehicle for thought and enables it to be carried immensely further, language is fitted into the thinking that originates in the first two years of life. If thought and language are to keep in step with one another as the child continuously models his language on that of adults, the child must be active in building his thought processes. If this is not the case, it is more likely that the child's verbalizations will lead adults to think that he understands more than he actually does. There is a grave danger of this happening between about 2 and 5 years of age since the child has copied the adult model of language yet he remains at the pre-operational level of thought. One fundamental lesson for parent and teacher from Piaget is that while language is important, mere verbalization and verbal knowledge are of little value in themselves.

4. Piaget (1926) indicated that social influences play a role in helping a child to move from cognitive egocentrism to systematic thought. Mere experience may have no effect on the available schemas, or an attempt may be made to distort the experience so that it could be assimilated. But social interaction with the peer group forces the child to re-examine his own thinking, so that he can satisfy his need to share the thoughts of others and his need to communicate with them. Argument with the peer group forces the child to reason with himself. Likewise, dialogue and discussion with parents and teachers, the answering of questions, etc., contribute to the growth of concrete operational thought.

5. Piaget's views clearly support what is broadly cal[illegible] a discovery approach to learning with a judicio[illegible] use of exposition at the elementary school level. [illegible]ut in high school more use can be made of exposition since the child's ability to appreciate the form of an argument improves, and he is able to elaborate concepts not themselves derivable from first hand experience. The Geneva school argues that the onset of the stage of formal thought is relative to the culture pattern. Beyond the age of 12 years or so, due to physiological factors, this level of thinking may be a product of the progressive acceleration of individual development under the influence of education and culture. The 13 year-old is not only dissatisfied with the gaps and uncertainties resulting from concrete operational thought,

but he is beginning to think beyond the present. This thinking ahead is likely to be determined by the experiences received at home, at school, and at work; his social attitudes; the climate of opinion and expectancy in the community, and the concepts that are frequently made use of in the society. In committing himself to the future, the adolescent begins to build theories.

All high school pupils thus need the opportunities to discuss with adults and teachers, and among themselves, viewpoints and theories relating to varied problems; the viewpoints and theories sometimes being in direct conflict with one another. They also need the chance to see, at first hand, the kinds of employment that will be available to them, and be in a position to think of, and discuss, their future roles as workers and citizens. They thereby get greater opportunities to commit themselves to possibilities. The greater the need to question and find out, to struggle for solutions to problems, and to commit oneself to possibilities, the greater seems the likelihood of formal thought developing. It may be that the culture pattern rather than the school plays the greater role in the growth of formal thought although we do not know if this is so in fact. Peluffo's (1964) study, which compared children born in Sardinia but who had been in Genoa for varying lengths of time, with children born and bred in Genoa, gives some indication of the likely effects of the culture pattern on the onset of formal thought.

6. In the upper-elementary classes and in high school, extrinsic motivation begins to influence the pupil. For example, he may become aware of what his parents or society expect of him; or he may fear failure or punishment if he does not work hard. One must not belittle this kind of motivation for it often stimulates the child to learn and to work as hard as he can at the level of thinking at which he is. But there is no evidence that it can, in itself, change the quality of the pupil's thought. Extrinsic motivation may well hold a pupil to a task, and in virtue of the prolonged child-task interaction enable him to restructure his own schemas. Thus external influences which are at work over a number of years, as, for example, parental encouragement, can play an important but indirect part in intellectual growth.

CONCLUSION

In conclusion, I would like to suggest areas in which we need research:

1. What are the long-term effects of very early stimulation of the culturally deprived and of certain types of school educable retarded children? If such is commenced during Piaget's sensori-motor stage of development, how do the long-term outcomes compare with those outcomes when the enriched environment is delayed until 3 and 5 years of age? The recently reported long-term study of Skeels (1965) in America suggests that very early stimulation may be of great importance. And what is the most suitable stimulation to provide at 6, 9, 15 months?

2. What are the long-term effects of an education based, as it were, on Piagetian ideas, particularly when they employ materials and situations of intrinsic interest to the child, from K through grade 4? Of course there will be some exposition, as there will be an increased proportion of time spent on exposition in the upper classes of the elementary school and in high school. What are the long-term effects of such methods over, say, 15 to 18 years? Is there any difference between such children and those brought up on other approaches at the second, fourth, eighth, and twelfth grades?

3. What is the effect of the emotional life on cognitive development? This is an important, although complex, question that cannot be brushed aside. In the 3 to 5 year period especially, fantasy - which I define as emotion clothed with images - is playing a role in the growth of thought. We ignore this at our peril. If in real life a child can dramatize his fantasy, he may free himself from the dominance of fantasy, and thought processes may develop smoothly. When the 2- to 5-year-old is presented with a story or situation which he only partly understands, he brings to it his experiences and his fantasies to fill in gaps and structure, say, the story. Each young child will have something of his own in the interpretation of the story, for information cannot yet be classified systematically and objectively. In the fairy tale particularly - which again is only half understood - the child's feelings of fear, anger, love, etc., which already exist, are perceived by him in the story. The story acts as a mirror and he is helped in recognizing the parts in himself in the "Good Prince" or "Bad Dragon." The "mirror-like properties" of fairy stories, and the fact that he can structure it with his own experience and fantasy, perhaps explains its abiding value. Similarly the child deals with his fantasy in painting or modeling. At 6 to 7 years of age fantasy is still at work, but the child is in better control. He may now accept a story that he could not tolerate at 4. I am no clinician in the accepted sense of the term, but I have enough experience with young children to know that the fantasy life is playing a role in the growth of thought processes although we have little evidence in a scientific sense on this point. What is the effect of kindergarten and school activities and which are likely to help the child to control his fantasy? What is the effect of using materials likely to have personal significance for him? The mathematician-logician may be forgiven for ignoring this when he is trying to teach kindergarten children something about sets. The psychologist and educator should realize that here is an important field about which we know little.

4. What is the effect of the culture pattern and subculture patterns on the growth of human thought processes? We know something of the relation between socioeconomic groupings and attainment but far less about such groupings and thought. Evidence from Hong Kong via Goodnow (1962), and from the Tiv tribe of Nigeria via Price-Williams (1961, 1962), suggests that culture pattern may not have as serious effect on the growth of concrete operational thought as on formal thought, especially if the materials used are well known to the testees. Nevertheless there is some delay at the level of concrete reasoning due to culture pattern as the work of Peluffo (1964) in Italy, and Almy (1966) in America. What are the characteristics of the sub-cultures that have the greatest downpulling effect?

5. We need to know why a schema, say, that of proportion cannot be applied to a wide variety of situations at the same time How important is familiarity with the problem, lack of specific experience in that area, and expectations based on previous experience?

6. We need a very great deal of information regarding the growth of more advanced concepts in mathematics and science, e. g., "function" and "entropy."

These questions that I have raised are broad ones, but ones of great importance. They are of significance to mankind, not to just the Americans or the British.

REFERENCES

Almy, M. Young children's thinking. New York: Teachers College Press, Columbia University. 1966.

Berko, J. The child's learning of English morphology. Word. 1958, 14. 150-177.

Berlyne, D. E. Structure and direction in thinking. London, Wiley, 1965.

Dodwell, P. C. Children's understanding of number and related concepts. Canadian Journal of Psychology. 1960, 14, 191-205.

Dodwell, P. C. Children's understanding of number concepts: Characteristics of an individual and of a group test. Canadian Journal of Psychology. 1961, 15, 29-36.

Dodwell, P. C. Relations between the understanding of the logic of classes and of cardinal number in children. Canadian Journal of Psychology, 1962, 16, 152-160.

Dodwell, P. C. Children's understanding of spatial concepts. Canadian Journal of Psychology. 1963, 17, 141-161.

Flavell, J. H. The developmental psychology of Jean Piaget. London: Van Nostrand, 1963.

Goodnow, Jacqueline J. A test of milieu effects with some of Piaget's tasks. Psychological Monographs, 1962, 76 (36, Whole No. 555).

Inhelder, B., and Piaget, J. The growth of logical thinking from childhood to adolescence. London: Routledge, 1958.

Jackson, S. The growth of logical thinking in normal and subnormal children. British Journal of Educational Psychology. 1965, 35, 255-258.

Lovell, K. A follow-up study of Inhelder and Piaget's The Growth of Logical Thinking. British Journal of Psychology, 1961, 52, 143-153.

Lovell, K., and Slater, A. The growth of the concept of time: A comparative study. Journal of Child Psychology and Psychiatry, 1960, 1, 179-190.

Peluffo, N. La nozione di conservazione del volume e le operazioni di combinazione come indice di sviluppo del pensiero operatorio in soggetti appartenenti ad ambienti fisici e socioculturali diversi. Rivista di Psicologia Sociale, 1964, 31, 99-132.

Piaget, J. The language and thought of the child. London: Kegan Paul, French, Trubner, 1926.

Piaget, J. Les stades du développment intellectual de l'enfant et de l'adolescent. In P. Osterrieth et al. (Eds.), Le problème des stades en psychologie de l'enfant. Paris: Presses Universitaires de France, 1956. Pp. 33-113.

Piaget, J. Introduction. In P. Gréco, J. B. Grize, S. Papert, and J. Piaget, Problèmes de la construction du nombre. Études d'épistemologie génétique. Vol. 11. Paris: Presses Universitaires de France, 1960. Pp. 1-3.

Price-Williams, D. R. A study concerning concepts of conservation of quantities among primitive children. Acta Psychologica, 1961, 18, 297-305.

Price-Williams, D. R. Abstract and concrete modes of classification in a primitive society. British Journal of Educational Psychology, 1962, 32, 50-61.

Ripple, R. E., and Rockcastle, V. N. (Eds.) Paiget rediscovered Ithaca: School of Education, Cornell University, 1964.

Skeels, H. M. Effects of adoption on children from institutions. Children, 1965, 12, 33-34.

Smedslund, J. Concrete reasoning: A study of intellectual development. Monographs of the Society for Research in Child Development, 1964, 29, (2, Whole No. 93).

Promoting Creative Thinking in the Classroom[1]

Martin V. Covington
University of California
Berkeley

THE LAST decade has seen a sharp increase in the volume of research on creativity and creative thinking with a major focus on isolating the basic processes of creative thought (Mednick, 1962; Mendelsohn and Griswold, 1964) and on identifying the unique characteristics of creative persons (MacKinnon, 1965; Barron, 1964; Crutchfield, 1963). Despite such unprecedented interest in creativity, until quite recently there has been relatively little research on the general problem of nurturing and promoting creative thinking, especially in the classroom setting. However, due partly to increased public awareness of the need to husband our intellectual resources, this important topic is beginning to receive the attention and research support it deserves.

The purpose of this paper is to describe part of the research program at Berkeley for developing curriculum programs aimed at promoting the general level of creative thinking among elementary-school children and to consider some of the pedagogical implications of such research, especially regarding the value, feasibility, and justification of teaching for creative functioning. For a survey of other current research in this area representing differing theoretical and methodological approaches, see Myers and Torrance (1964), Osborn (1953), Parnes (1965), Suchman (1961), and Upton and Sampson (1961).

By way of introduction, it should be noted that most of the current attempts to develop such teaching materials, including that of the Berkeley group, share two fundamental assumptions: (1) that all children, regardless of age or initial intellectual level, fall far short of realizing their potential for creative thought, and (2) that these individuals can be taught to make fuller and more appropriate use of their capacities by means of systematic and direct training of certain cognitive skills.

As for a definition of a creative child (in a real sense the ultimate criterion by which the success of a teaching program must be judged), it is generally agreed that he is intellectually curious and exploratory, taking an active part in manipulating and reconstructing his environment. To paraphrase a number of current definitions:

> The creative child prefers to explore the unknown, rather than to conserve the already known; he prefers explaining facts in new ways, rather than continuing to rely on traditional well-established explanations; he indulges in adventuresome thinking and raises questions, rather than being content with things as they are.

With these preliminary comments setting the stage, we turn next to more pointed observations about the nature of creative thinking and how best to foster it — observations which to a large degree have determined the specific form and character of the research of the Berkeley group. As a first step, we will identify and discuss several widely-held misconceptions about creative functioning and its facilitation.

MISCONCEPTIONS CONCERNING THE NATURE OF CREATIVE THINKING

Perhaps the most commonly held belief concerning creative functioning is that the child will become creative of his own accord – by drawing on some heretofore latent potential – if only he can be placed in a stimulating yet permissive and nurturing environment. The erroneousness of this view is due not so much to a misreading of the nature of man as it is to an oversimplification of the means by which his latent capacities are to be realized. More specifically, an environment which is at once permissive and stimulating undoubtedly plays an important role in establishing the conditions necessary for creative output, but this is only part of the story. In order to take full advantage of an unrestrictive atmosphere, the child must first come to understand what constitutes creative ideas in the given situation and how he can achieve such ideas for himself. In short, he must learn how to think creatively. In this connection, a growing body of observational and empirical evidence (Hammer, 1961; Holt, 1964; Peel, 1960; Torrance, 1965) indicates that most school children are inadequately prepared for creative thinking. Corroborative data gathered by the Berkeley group at the fifth-grade and sixth-grade levels reveal a conspicuous inability among students to think of any ideas, much less clever or novel ones. Moreover, there is little understanding of the nature of originality and virtually no sense of planfulness which is necessary for prolonged effective work on a creative task. Not surprisingly then, one finds that the responses of most children to challenging tasks, even in a permissive environment, are by and large pedestrian and generally lacking in creative merit.

In order to think creatively the child must have at his disposal a repertoire of creative thinking skills. To name a few: the ability to recognize gaps in existing information; a facility at formulating relevant questions, and a sensitivity to the demands of the task so one can adequately judge the suitability of proposed ideas. The epitomy of creative thought is a sense of "disciplined abandonment." An array of cognitive skills such as those just mentioned constitutes the "discipline" component which must be present in concert with an accepting nurturing environment before the child can take full advantage of such freedom and "abandon" himself to the task at hand. Thus, seen in this perspective, the job of fostering creative thinking is a more complicated undertaking than suspected by many. It is more than merely providing a permissive atmosphere for the release of creative potential; it is more like teaching the child how to make use of his freedom to create.

If one must teach children to think creatively, then what kinds of pedagogical strategies would seem most appropriate? There are, no doubt, a number of reasonable answers to this question. For example, one approach – the traditional one – is to teach in such a manner that the child will come to act creatively in a number of different subject-matter disciplines. Another contrasting, yet complementary possibility is to teach for a number of cognitive skills fundamental to all creative thinking and then show the student how such generalized skills can be applied in specific subject matter areas. This latter strategy carries with it the implication of developing curriculum materials whose subject-matter is the creative thought process itself. As we shall see presently, it is this approach which is being explored by the Berkeley group.

Although there is considerable latitude with respect to the variety of reasonable approaches for promoting creative thinking in the classroom, there are limits. Some limits can be established based on our growing knowledge of the dynamics of the creative process. For example, it has been well established that creative thinking is largely dependent on cognitive operations and dispositions of a non-logical character. Yet despite this, the notion as well as the practice persists of teaching for logic and for critical thinking as means for promoting creative functioning. As we shall see shortly, creative thinking, whether it be in science or in the humanities, involves a good deal more than rational and critical analysis, and to teach only for these functions to the exclusion of more imaginal speculative processes is to fall short of fostering creative thought.

It has been further argued that teaching for scientific methods of thinking will establish a respect for the lawfulness and the simplicity of nature itself. Although the child may indeed infer from such exercises that nature is orderly and lawful, such understanding may curiously enough act to fetter creative thought. More specifically, as currently taught, elementary school science, the humanities, and the social sciences as well, do not always reflect the way scientific investigation and creative scholarship actually happen. Creative contributions to any field of knowledge are made by wrestling with ambiguous conflicting facts and data with the overall aim of bringing conceptual order out of chaos. However, as represented in many current textbooks, the drama of the pursuit of such order and understanding is replaced by an oversimplified, tailored, and predigested view of a discipline, emphasizing what is already known and stressing the all-encompassing explanatory power of presently accepted theories or laws. Such a simple, clear-cut presentation which smooths over complexity and controversy is designed primarily to make for more efficient learning and retention, rather than to focus on the problem of how the child himself can learn to manipulate a given set of facts, to draw his own conclusions, or indeed to discover new facts, which is after all, one of the primary functions of the creative innovator. There is little question that such textbook presentations will make for the most rapid, untroubled assimilation and mastery of the material, but it should not be assumed that the approach which makes for the most efficient learning will also be most effective in fostering creative and productive thinking.

Moreover, in the way most subject matter has been traditionally taught, the student is rarely allowed to practice on problems which require innovative modes of thought for their solution. Consequently, he has no direct opportunity to learn what constitutes creative ways to manipulate data, nor to develop an intuitive sense for designing research, nor to ask questions which favor serendipity and

maximize the chance for new insights. In short, the child does not acquire the cognitive skills which would be most helpful in preparing him for future productivity.

A specific case in point illustrating the lack of emphasis on thinking creatively about facts and information is the set of so-called "discussion questions" found at the end of many textbook chapters. As commonly used, these questions are not particularly helpful in promoting creative patterns of thought. Often the student is merely asked to think about these questions with little or no subsequent follow-up discussion of the adequacy of his answers. With minimal dialogue between the student and the teaching authority, the child rarely gets any systematic feedback as to what counts as merely an adequate answer as opposed to an ingenious, novel one. Even in those cases where the teacher finds time to comment on the quality of the pupil's essays, the feedback is inevitably delayed, thus reducing the effectiveness of the exercise. Because the child typically operates from question to question in what is essentially a "cognitive vacuum" with little opportunity to profit from his past efforts, he is unable to develop any internal standards of excellence against which he can evaluate subsequent ideas. In effect this means he is unable to modify his efforts in the direction of more creative and imaginative output.

Another feature of such discussion questions – their placement at the end of the chapter – strengthens the impression that these exercises are simply an appendage, added to the text as an afterthought rather than forming a central part of the learning experience itself. Moreover, their placement occurs at the psychologically inappropriate moment for stimulating productive ideas. By the end of the chapter much of the intellectual tension and drama built up as part of the content presentation will have dissipated, making it difficult to engage further the child's interest. In contrast, the placement of discussion questions should be coordinated with the presentation of the content itself, capitalizing on the student's immediate curiosity and momentarily aroused interest.

RECOMMENDATIONS FOR A CURRICULUM OF CREATIVE THINKING

Recognizing some of these pedagogical deficiencies, what can be recommended with regard to a curriculum for creative thought? The first recommendation is that such a curriculum feature repeated opportunities for the child to practice wrestling with complex challenging tasks of the kind that might, in a more complex form, engage the efforts of creative adults. Of course, such problems (which we shall call "creative-tasks-in-miniature") must be scaled down and simplified in many respects – in particular, reducing their dependence on technical or specific knowledge. Nevertheless, even with various modifications of this type such tasks could still retain the basic elements of any intellectually creative undertaking. A second recommendation is that certain broad guides and strategies for creative thinking be introduced as the child works on these tasks. Such guides would act as points of reference around which the child could organize his work and would impart an atmosphere of planfulness to the process of problem solution.

Next, what can be said more specifically about the nature of these "creative-tasks-in miniature" – their particular content and the composition of the accompanying rules and strategies for productive thought? It is sobering to realize that the creative innovators of the year 2000 (only 32 years hence) are to be found in the elementary schools of today. As these individuals stand on the threshold of the twenty-first century, they will be confronted by scientific revelations and social issues literally undreamt of today as well as a large share of old problems left unsolved by previous generations. To meet these challenges the innovator of tomorrow must create new social systems, new scientific techniques, tools, and theoretical structures. To accomplish this, he cannot rely on presently fashionable views as to what constitutes fruitful scholarship or good scientific method, nor can we afford to instill in him a blind devotion to currently fertile theories, since much of this theoretical apparatus is already approaching obsolescence. Here then is a compelling argument for the necessity of teaching broad strategies for creative thought and of encouraging cognitive flexibility and intellectual curiosity, rather than merely teaching for what we presently take to be true or useful.

What kinds of tasks could children practice on now to prepare them for their roles as the innovators of the future? Two kinds are suggested: the hypothetical problem and the enduring problem. The former type minimizes reliance on present facts and theories in obtaining a solution, while the latter type focuses on the deathless issues of human affairs which are certain to be the subject of creative reinterpretation and re-evaluation in any future age.

Typically in the case of the hypothetical problem certain artificial conditions are postulated which do not correspond to present realities, and the individual is called on to operate within these internally consistent but artificial systems. One problem of this type, which itself has become popular as a measure of creativity, is the consequence test: given certain conditions, such as a world in which time runs backwards, what are all the consequences? Such problems demand cognitive flexibility, imagination, and the ability to adopt as temporarily real a set of foreign and unfamiliar assumptions.

At Berkeley we are now drawing up plans for an experimental teaching apparatus which simulates a visit to a strange planet. The student is put in charge of the first exploratory expedition. His job, a highly creative and demanding one, is to come to understand and to explain the welter of strange events which he will encounter in this alien, unfamiliar world where the usual physical, biological, and social laws are suspended or greatly modified. Here the child is literally confronted with a hypothetical world which, although it is ordered and internally consistent, operates in ways that run counter to his previous experience. He must suspend his present expectancies about the world as he knows it long enough to create new explanations for strange and perplexing phenomena. Thus, by permitting the child to deal with hypothetical tasks which

call for the creation of new systems of ideas, independent of present facts and assumptions, he can hedge against the time when these facts will outlive their usefulness and will no longer serve as a basis for productive thought.

Granting the largely unknown course of future events, there are nevertheless certain enduring problems of man and society which are constantly reasserting themselves and no doubt will engage the attention of future creative thinkers. The young student of today can best be prepared to cope with such issues in adulthood by being presented now with repeated opportunities to deal with enduring themes and quandries: to search for new implications and to re-phrase them in contemporary terms. Many of the most suitable teaching examples are found in literature, such as the Greek myth of Sisyphus, dealing with man's reluctance to accept his fate; the tragedy of Frankenstein, reflecting the consequences of man's age-old dream of creating a more perfect being is his own image; or the mythical Icarus, personifying man's unflagging zeal to explore nature regardless of the consequences. Of course, it is understood that such issues cast in the form of teaching examples would have to be greatly simplified and made appropriate to a child's level of understanding. Notice, however, that such modifications are made easier by the fact that the traditional vehicles for the dramatization of these issues are themselves reasonably constant - the novel, the play, and the poem - providing already familiar media through which children can express their ideas. In this connection, we shall presently describe a "creative-task-in-miniature" in which the student creates a play dealing with yet another timeless and recurring theme - the plight of the wrongfully accused.

Finally, what are the sources on which one can draw in formulating a set of broad cognitive skills and strategies which could be taught now to prepare children for future productivity? Information concerning such generalized cognitive strategies can be gleaned not only from laboratory research dealing with complex problem solving and thinking but from anecdotal and life-history accounts of highly creative individuals as well. In the first instance there is a steadily accumulating body of knowledge from the psychological laboratory concerning the kinds of factors, both facilitative and inhibitory, which influence the course and quality of productive thought. One widely documented example is the deleterious effect of rigid mental set on effective problem solving (Luchins, 1942). Here, because the individual initially formulates or perceives a task in a narrow and inherently biasing way, he is effectively sealed off from attaining a solution. A number of appropriate thinking strategies incorporating these and other research findings can be formulated. For example, with reference to the present example of mental set, the student can be taught to view a problem broadly, to re-phrase questions in new ways, and to resist jumping to premature conclusions.

The other source of cognitive strategies comes from anecdotal descriptions of creative individuals at work (Ghiselin, 1952; Koestler, 1964). A number of insights into the fundamental nature of creative thought have grown out of such observations as well as the discovery of important phenomena associated with creativity, such as incubation and intuition. As an example, the famous and widely cited self-reports of Poincare, the French mathematician, have served to focus attention on the conceptually fertile notion of creativity as basically a process of recombining ideas in new and novel patterns. And of most relevance to our immediate concern, such observations have led to the elucidation of a number of specific problem-solving strategies, among them the technique of identifying the essential elements of a problem situation and deliberately juxtaposing them in various ways (Campbell, 1960).

By combining the experimental laboratory approach and its virtues of empirical validation with the rich and broadly inclusive self-reports of the highly creative thinker, one can formulate a set of rules and strategies which are pervasive enough to apply to a variety of creative tasks, yet simple enough to be readily understood by elementary-school children.

In summary of the first section of this paper it has been argued that before the student can take full advantage of a permissive stimulating atmosphere he must be taught how to think creatively. In developing a teaching program for creative thought the single most important pedagogical issue is the long-term social and personal usefulness of what is taught. It has been suggested that one reasonable teaching strategy, in light of this consideration, is to strengthen those cognitive skills basic to all creative thought in the context of complex yet meaningful problems which reflect the principal steps in the creative act. These observations, taken as a whole, have formed the broad guidelines for the research carried out by the Berkeley group over the past several years.

THE PRESENT RESEARCH

As a first step it was decided to develop a set of prototype instructional materials dealing with only one aspect of productive thought- creative problem-solving. If the particular teaching approach showed promise, then other programs would be designed focusing on yet other domains of creative thought such as creative innovation and creative expression.

After several years of intensive effort involving numerous pilot tryouts, successive revisions, and several full-scale experimental studies, the General Problem Solving Program was developed (Covington, Crutchfield, and Davies, 1966). The General Problem Solving Program (GPSP) consists of a series of sixteen "creative-tasks-in-miniature" designed for the fifth-grade and sixth-grade levels. These problem episodes act as a vehicle by which the student practices a number of broad rules and strategies concerned with various facets of effective problem solving, such as stating the problem in an open and unbiased fashion, reformulating the problem in terms of familiar metaphors and analogies, or developing a systematic plan for generating ideas.

Each lesson presents a mysterious occurrence or unexplicable happening which the student tries

to explain, such as the puzzling behavior of a group of aquanauts during a deep-sea dive. This general theme – the mysterious occurrence – was chosen because it encompasses in a natural and uncontrived manner many of the fundamental aspects of the problem-solving process. In addition, such a theme is neutral with respect to orthodox curriculum materials. This makes for a broader transfer of what is taught than might otherwise be the case if the principles and concepts had been associated primarily with a specific content area. Finally, children exhibit a keen interest in working on problem situations featuring a mystery-detection theme.

Each lesson is presented in booklet form and is cast in a self-instructional linear format. (For a theoretical discussion of the potential uses of programmed instruction for fostering higher-order thought processes, see Crutchfield and Covington, 1965.) Briefly, the child works through the lesson in a step-by-step sequence by himself and at his own pace. As the problem unfolds page-by-page, with the advent of new facts and clues, the student is called on to reformulate the problem in his own words, to list questions he considers crucial to the solution or to devise a plan of action which he intends to follow. The student receives immediate guidance in his efforts on successive pages of the booklet. This guidance typically takes the form of presenting a range of ideas, questions or courses of action which the student might have thought of in the given situation. Here the emphasis is on encouraging appropriate diversity in the student's thoughts by showing him within wide limits what constitutes valuable ideas, crucial questions and fruitful ways to formulate problems.

The GPSP is also designed to promote beneficial attitudes toward productive thinking. One of the basic strategies is to increase the child's experience in coping successfully with thought problems. Each problem episode is sequenced so that as the student works through the booklet he is exposed progressively to more clues and hints. Thus, each child, independent of his initial capacities or personal reservations about thinking, will at some point come to discover the solution for himself. In addition there is a systematic attempt to foster, by means of appropriate teaching examples, a number of attitudes which favor effective problem solving, such as open-mindedness, persistence, and suspension of premature judgment.

A story-line is maintained throughout the GPSP to supply a sense of continuity, and perhaps most importantly, to provide the reader with a set of identification models. The story concerns two school children, Jim and Lila (brother and sister), whose extra-curricular pastime is the exploration of mysterious situations which arise in and around their hometown. Jim and Lila are assisted – and when the occasion demands, assiduously directed step-by-step – through these adventures by their uncle. The uncle in addition to being a high-school science teacher also "moonlights" as a detective.

The identification-model technique is intended to introduce the reader gradually to the difficult and often frustrating process of becoming a more effective thinker. For example, the roles of Jim and Lila reflect the vicissitudes associated with all complex problem solving – the long discouraging periods of intense effort with little apparent progress, the inevitable setbacks, and finally, the intellectual elation of discovering a solution. Through Jim and Lila the reader can experience all this vicariously without being plunged immediately into the situation. After the first few lessons, however, the reader is gradually drawn into the various problem-solving activities by being requested to think of and to record his own ideas in concert with the efforts of Jim and Lila – first the student generating his own ideas or questions, then Jim and Lila responding with theirs. The feedback examples provided for the reader are presented as Jim's and Lila's ideas.

Against this backdrop of transitory day-to-day successes and failures, a long-range change takes place in Jim and Lila. They are depicted as overcoming initial handicaps of reticence, apathy, and negativism toward thinking until ultimately, but not without an occasional setback, they become reasonably comfortable with their own thought processes, much more enthusiastic about tackling problem situations, and more confident in their own abilities to cope with them. It is hoped that this subtle but perceptible change in Jim and Lila will foster in the reader a sense of his own progressive improvement. At the same time it is intended that as many readers as possible will finally come to surpass Jim and Lila in problem-solving proficiency and will outgrow any earlier dependence on them as a source of ideas and inspiration.

The inclusion of the uncle provides Jim and Lila, and presumably the reader as well, with a benevolent authority figure and confidant who not only nurtures and encourages the nascent attempts of these tyro-thinkers but who also shows them how to think for themselves by means of appropriate rules and strategies. Additionally, the uncle stands as a valued model personifying an enthusiasm for intellectual exploration.

Each lesson is presented in an illustrated format. Such a dominantly visual presentation not only increases the student's interest in the materials, but makes it easier for him to follow the necessarily complicated dialogue and thought sequences of the story characters. The lessons contain an average of 40 pages and are designed to be presented at the rate of one lesson per day. The average time taken to complete a lesson is approximately 35 minutes.

Several separate studies (summarized in Covington and Crutchfield, 1965) have been carried out using various preliminary editions of the GPSP which to date have involved a total of some 481 fifth-grade and sixth-grade school children from the San Francisco Bay Area, of whom 267 were given the training materials and the remaining 214 were used as controls. The basic design of these studies is essentially the same. All students are first administered an extensive pretest battery (six hours in length in the latest study) which consists of a number of tests of creative thinking (Torrance, 1965), tests of problem-solving ability (Covington, 1966c), inventories designed to measure the child's attitudes toward problem solving and thinking (Olton, et. al., 1967), and various tests of school achievement and

scholastic ability. These data are used as a basis for the initial matching of the classrooms. Individual classrooms are matched in pairs; one classroom from the pair is then assigned at random to the instructed condition and the other to a control condition. In each of the various studies to date, the instructed groups have always been administered the GPSP. In contrast, a number of different control conditions have been used to test various hypotheses about the nature of the training effect. The typical case, however, and the one on which the present analysis is based, calls for the administration of a similar but shorter set of self-instructional materials whose content is unrelated to problem solving. The purpose of the control program is to insure a sense of involvement and participation on the part of the control children and to make certain that they are well acquainted with the self-instructional format. Following the training period all classes are administered a posttest battery which includes a repeat of the attitude inventories and the tests of problem-solving ability along with parallel forms of the tests of creative thinking.

Several examples will suffice to illustrate the general types of tests included in the pretest and posttest batteries – their content, style and degree of similarity to the training material. One type reflects the more traditional tests of problem solving. In these instances, because of a large number of restrictions placed on the task, there is only a limited set of answers or principal solutions which will satisfy the problem in its entirety. Here an emphasis is placed both on convergent and divergent thinking. The student is called on to diverge in his search for appropriate ideas – opening up a host of possibilities – but at the same time to converge, focusing on the few most suitable ideas. A representative sample is the X-ray problem (a modification of the classic problem used by Karl Duncker). The child is required to invent a method to kill a tumor deep inside a body by using an X-ray but without harming the surrounding healthy tissue. The primary constraint is that if the X-ray is too strong it will kill both the healthy tissue and the tumor. If, on the other hand, the X-ray is made too weak, it will not harm the good tissue, but neither will it kill the tumor. Incidentally, this problem in addition to possessing a high degree of curriculum relevance is quite unlike anything encountered in the training lessons in terms of content or theme.

The other main type of problem included in the criterion batteries is the Minnesota Tests of Creative Thinking (Torrance, 1965). Such problems are quite open-ended with few constraints. This allows for a number of solution-ideas, no one of which can be judged as the most suitable. One example featuring a product-improvement theme requires the child to think of all the ingenious clever ways to make a toy dog more fun to play with.

As to the results: in general, the outcomes of the various studies employing the GPSP have been highly consistent. The performance of the instructed children is markedly superior to that of the control children both on the tests of problem-solving ability and on the tests of creative thinking. (For a detailed report of findings as well as a discussion of scoring procedures, see Olton, et. el., 1967.)

In order to gain a clearer picture of the over-all magnitude of this training effect, consider several performance indices such as total ideas generated, total quality of ideas, mean quality per idea and the incidence of principal solutions. Composite scores for each of these indices are computed for every student by summing over his performance on a number of problem-solving tests in the posttest battery. In the same manner a similar set of composite scores is obtained for the tests of creative thinking. On comparing the mean values for the instructed and control groups on any of these composite indices it is found that the instructed groups are invariably superior. In most cases the magnitude of these differences is absolutely large and beyond mere statistical significance. This is illustrated by the fact that when comparing frequency distributions for each composite index, the fiftieth percentile of the instructed groups typically falls around the seventy-fifth percentile for the control groups.

Another way to gauge the degree of educational significance of a training effect is to compare the level of performance of instructed and control children of differing IQ levels. In this connection it has been found (Covington, 1965; Covington, 1966b) that the mean performance of instructed children with IQ's below 99 (mean IQ = 91) is on a par with the mean performance of control children whose IQ's fall between 100 and 115 (mean IQ = 107). This indicates that the administration of the GPSP makes for a substantial boost in the performance of low IQ children over a wide variety of test problems, many of which are curriculum relevant.

To determine how long these training effects persist, a follow-up test battery was given five months after the administration of the posttest in two different studies. These follow-ups included approximately 80 percent of the fifth-grade children from the original samples who by this time were in sixth-grade classrooms with new teachers. Wherever possible every precaution was taken to insure that the students did not recognize the tests as a part of the earlier work. For example, the children recorded their ideas on regular school paper, rather than in a standard booklet of the kind used in the original study. All the tests were administered by the teacher herself, and whenever possible were introduced as part of the regular classroom work.

In the case of tests of problem-solving ability, the instructed children continued to surpass their control counterparts on the various performance indices, and although the margin of differences had diminished somewhat as compared to the magnitude of effects found at the time of the original posttest, the differences were nevertheless found to be statistically significant. The picture is not as consistent for the tests of creative thinking. In one sample with a total of 108 children, the instructed group was superior to the control group, whereas in another slightly larger sample these differences washed out. It appears that after a five-month interval the training effects are at best marginal in the case of creative thinking tests. This is not surprising when it is realized that as a group these tests represent a type of problem not directly trained for in the GPSP. Consequently they would be more likely to reflect a

greater diminution in training effect than would the tests of problem-solving ability which are more directly amenable to the kinds of strategies taught for originally.

ISSUES OF VALUE AND JUSTIFICATION

One of the most fundamental questions raised by the present research is whether programs designed to promote creative thinking are actually needed, seen in the perspective of a child's total educational career. No matter how beneficial a given teaching program may be, it is likely that untutored children (the so-called controls) will sooner or later catch up, simply through the normal process of intellectual maturation and accumulating experience. In the present data the sizable reduction in the magnitude of the training effect five months after the administration of the GPSP supports this contention. Of course, it can be argued that more permanent – and in this sense more meaningful – changes will occur only on an accumulating basis, and that what is needed is a program covering a whole school year or even longer, consisting of interlocking, coordinated curriculum units which build on one another. It remains a moot point, without empirical support, yet it would seem likely that a long-range program of increasing scope and complexity designed to stimulate intellectual growth would allow the instructed child not only to maintain a performance superiority, but actually to increase that margin as the program proceeds. While it is true that the trained and the untrained student alike enjoy a natural expansion of intellectual capacities, there is grave doubt that the untutored children would ever make use of their burgeoning capabilities to the same extent as would children who received systematic long-term training. And, as we have seen, since most elementary-school children make scant use of their present capacities for creative thought, there is little reason to believe that these same children at a later age would spontaneously draw on their capacities to any greater degree.

A related point concerns the developmental changes in attitudes and values which favor the exercise of creative thought. While it is quite clear that sheer proficiency in various cognitive skills such as question asking or idea generation is a function of age, there is as yet no evidence that the relevant attitudes and values increase as the child grows older. Thus, even though untutored children may in time overtake the trained children in terms of sheer proficiency, they may be markedly deficient in the very attitudinal dispositions necessary to put such skills to meaningful use. In this connection, creativity training can be designed to provide the student with experiences he is unlikely to receive anywhere else – the challenge of working on a complex but meaningful problem or the sense of satisfaction at discovering a solution – experiences which may affect the child's emerging set of values and attitudes regarding creating functioning. Such experiences, if they occur at decisive points in the individual's development, especially in childhood, may be sufficient to bring him permanently past a critical threshold for actualization of his creative potential. As a matter of fact, we may find in the last analysis that the key to the unlocking of creative potential is to strengthen sets of attitudes and values which predispose the student to undertake creative tasks in the first place, rather than to start by increasing his sheer competence for performance by means of "skill training."

FUTURE RESEARCH PLANS

Research plans for the immediate future take two main directions. The first of these involves further development and expansion of the GPSP. In one instance a set of supplementary exercises is being developed to provide additional practice on the various skills and strategies taught in the lessons. These exercises are primarily for remedial purposes and can be used when needed, depending on the rate of progress of the individual learner. Another project concerns the development of a set of "curriculum-link-units," that is, lessons designed to illustrate how the student can apply the skills and strategies taught for in the GPSP to his actual school work in science and the social studies.

The other main direction of research involves the development of new self-instructional teaching units which introduce the student to domains of creative thinking other than problem solving. One obviously important area is that of creative understanding. Here, typically, one is confronted with a series of complex events or occurrences which in the initial phases of investigation often appear to be unrelated. The task is one of discovering meaningful relationships between such events and in some instances of predicting how these events would covary under various conditions. In the case of problem solving defined in a more traditional sense, the search is typically of a more limited nature – that of discovering a single workable solution-idea which satisfies a certain requirement, but often without the necessity of understanding why it works. Such a pragmatic emphasis tends to make for closure once a solution is obtained. On the other hand, creative understanding not only encompasses the discovery of workable ideas, but more often involves an ongoing procedure of checking and re-checking the validity of the proposed solutions against an accumulating set of facts and information.

Another area of creative output, perhaps the one most generally assumed when reference is made to creativity, is that of creative innovation. Here the individual invents or creates new problems, new systems of thought, or new products which did not exist previously. One example of a prototype training unit presently being developed involves the creation of a school play. The story-line concerns some malicious damage done to a barn owned by a local farmer and the farmer's unwarrented accusation that a certain boy is responsible. The boy, who is innocent, is able to convince his classmates of this fact, but is unable to prove it to a degree that satisfies the farmer and the rest of the adult community. Understandably the children react with indignation, but at the same time they realize their essential helplessness at being unable to defend adequately their interests in an adult world. While some of the children make plans to track down the real culprit, others decide to express the group feeling of ineffectuality and frustration by means of a play. Parenthetically, such a plot can be used to illustrate that there are other alternatives to the expression of

one's emotions besides direct aggression and that intense feelings can serve as a basis for the development of useful and satisfying products. The particular theme - the plight of the wrongfully accused - was chosen because it sounds a note of high drama and arouses immediate interest and sympathy in children who typically have a keenly developed sense of fair play. Moreover it represents - even when placed in such a simplified context - one of the enduring issues of man and society which demands new solutions from each succeeding generation.

The same general pedagogical devices are used here as in the GPSP. The student is led to develop the play in a step-by-step sequence, while thinking of his own ideas and being provided in turn with appropriate feedback and redirection of his efforts. In the early lessons the child works through sequences dealing with the creation of the basic plot. The student is encouraged with the assistance of identification models in the story to generate ideas which seize on the malicious incident as a point of departure from which to explore the more general implications of being wrongfully accused. This is in contrast to the more pedestrian and common-place approaches of simply dramatizing the boy's innocence or holding the farmer up to ridicule. Basic to the success of this teaching sequence is the introduction of metaphorical and analogical modes of thinking. Once such concepts are grasped by the student he is able to go far beyond a strictly literal interpretation of a situation and can generate a virtually limitless set of implications.

After the outlines of a clever insightful plot are secured, the student explores a number of detailed matters, all of which are important to the final product. In one lesson he creates ideas for the most effective scenery and staging, given a limited amount of material to work with; in another he is called on to create segments of the dialogue. The last lesson in the series simulates the public reaction to the play which represents a final confirmation of all the preceding activities. Various members of the audience including the farmer and the parents of the accused boy react to the play. There is general recognition that it was worthwhile in its own right as well as being helpful in clearing up a community misunderstanding.

FINAL PARADOXES AND DILEMMAS

Inherent in any attempt to teach for creative functioning are a number of dilemmas and paradoxes which give rise to both methodological and theoretical difficulties. In conclusion we will consider one recurring dilemma, not because it is necessarily representative or even the most fundamental, but rather because it indicates something of the extraordinary challenges encountered in developing teaching materials for fostering complex cognitive processes. Basic to the act of teaching is a guidance function. To a greater or lesser degree the student's behavior is necessarily guided, shaped, and redirected by a teaching authority. The student never has complete freedom. He is exposed only to certain learning experiences, in certain teacher-determined sequences. He is encouraged to talk about these experiences in only certain ways, using prearranged kinds of terminology. If we teach for creative functioning, then the child will inevitably be subject to some form of guidance. To the extent that an educational theory stresses the "discipline component" of creative thought, any forthcoming curricula will be more or less restrictive and predetermined. How then can we reconcile these procrustean-like features of teaching with the fact that in the last analysis creativity is primarily characterized by imaginal freedom and a spontaneous individualization of thought? Perhaps the answer, most simply put, is to teach for structured spontaneity. One technique adopted for this purpose and employed in our current work is to give the student direct practice in discriminating among ideas which are both unique and appropriate to a given task and ideas which are merely bizzare. In such a fashion the student learns to limit the reaches of his spontaneity. Another technique which requires a relatively long term training program involves comprehensive guidance in the early phases of learning. But, as the child becomes more proficient in the various skills and strategies, the rigid guidance of the program is gradually reduced. In this manner the student comes ultimately to rely on his own resources and initiative with only occasional redirection from the program. As the student becomes more self-sufficient he can experience yet another dimension of intellectual independence - that of determining for himself which of several creative tasks he will work on from among a number of alternatives. For example, in the case of the creative drama unit just described, the child could be given the option of either writing dialogue or of developing the scenery and staging, once he had completed the introductory units.

All the foregoing implies a complex catering to the individual differences of the learner; first, an intensive close-knit and personalized guidance of the child in the early phases of learning and later, encouragement of self-direction and self-determination. In both cases the training environment must be capable of selecting, sequencing and coordinating a great many combinations of material. This immediately suggests computer-assisted teaching. The self-instructional format of the current and proposed teaching programs is admirably suited to a computer-assisted operation. Indeed, another facet of the work of the Berkeley group is that of adapting the GPSP for a computer presentation, thus by-passing the cumbersome booklet format with its limited feedback capabilities. With computer assistance the child can be guided through a much more individualized sequence of learning, dictated by his initial level of competency, his particular cognitive style, and his individual rate of progress. Conceptually, it may appear that computers and programmed instruction are antithetical, perhaps even perversely so, to the task of promoting creative thinking. However, such an unlikely juxtaposition points to the kinds of imaginative responses which the educational community must make to the challenge of educating individuals in an increasingly complex and demanding world.

FOOTNOTE

1. The research reported in this paper was supported by a grant from the Carnegie Corporation of New York under the direction of the author and Dr. Richard S. Crutchfield. This paper

covers research conducted through 19ʟ'. Summaries of our more recent research activities can be found in Olton, et. al., (1967) and Olton and Crutchfield, in press.

REFERENCES

Barron, F. The relationship of ego diffusion to creative perception. In C. W. Taylor (Editor), Widening horizons in creativity. New York: Wiley, 1964. Pp 80-86.

Campbell, D. T. Blind variation and selective retention in creative thought as in other knowledge processes. Psychological Review, 1960, 67, 380-400.

Covington, M. V. The effectiveness of training for problem-solving proficiency and creative thinking as a function of differing ability levels among children. Paper presented at the meeting of the Western Psychological Association, Honolulu, June 1965.

Covington, M. V. A childhood attitude inventory for problem solving. Journal of Education Measurement, 1966, 3, 234. (a)

Covington, M. V. Programmed instruction and the intellectually gifted: Experiments in productive thinking. Paper presented at the Symposium on Programmed Instruction and the Gifted Child, International Convention of the Council for Exceptional Children, Toronto, April 1966. (b)

Covington, M. V., and Crutchfield, R. S. Experiments in the use of programmed instruction for the facilitation of creative problem solving. Programmed Instruction, 1965, 4.

Covington, M. V., Crutchfield, R. S., and Davies, Lillian B. The Productive Thinking Program. Series One: General problem solving. Berkeley: Brazelton Printing Company, 1966.

Crutchfield, R. S. Independent thought in a conformist world. In S. M. Farber and R. H. L. Wilson (Editors), Control of the mind. Part 2. Man and civilization: Conflict and creativity. New York: McGraw-Hill, 1963. Pp. 208-228.

Crutchfield, R. S., and Covington, M. V. Programmed instruction and creativity. Programmed Instruction, 1965, 4.

Duncker, D. On problem-solving. Translation by Lynne S. Lees. Psychological Monographs, 1945, 58, (5, Whole No. 270).

Ghiselin, B. The creative process: A symposium. New York: New American Library, 1952.

Hammer, E. F. Creativity: An exploratory investigation of the personalities of gifted adolescent artists. New York: Random House, 1961.

Holt, J. How children fail. New York: Pitman, 1964.

Koestler, A. The creative act. New York: Macmillan, 1964.

Luchins A. S. Mechanization in problem solving: The effect of Einstellung. Psychological Monographs, 1942, 54, (6, Whole No. 248).

MacKinnon, D. W. Personality correlates of creativity. In Mary Jane Aschner and C. E. Bish (Editors), Productive thinking in education. Washington, D. C.: National Educational Association, 1965. Pp. 159-171.

Mednick, S. A. The associative basis of the creative process. Psychological Review, 1962, 69, 220-232.

Mendelsohn, G. A., and Griswold, Barbara B. Differential use of incidental stimuli in problem solving as a function of creativity. Journal of Abnormal and Social Psychology, 1964, 68, 431-436.

Myers, R. E., and Tor   e, E. P. Invitations to thinking and doing. New York: Ginn, 1964.

Olton, R. M., Wardrop, J., Covington, M. V., Goodwin, W., Crutchfield, R. S., Klausmeier, H. J., Ronda, T. The development of productive thinking skills in fifth-grade children. Technical Report, Research and Development Center for Cognitive Learning. Madison, Wisconsin: The University of Wisconsin, 1967.

Olton, R. M., and Crutchfield, R. S. Developing the skills of productive thinking. To appear in: P. Mussen, J. Langer and M. V. Covington (Editors), New Directions in Developmental Psychology. New York: Holt, Rinehart and Winston, in press.

Osborn, A. F. Applied imagination. New York: Scribner's, 1953.

Parnes, S. J. A self-instructional course in creative problem solving. Unpublished manuscript, State University of New York at Buffalo, 1965.

Peel, E. A. The pupil's thinking. London: Oldbourne, 1960.

Suchman, J. R. Inquiry training: Building skills for autonomous discovery. Merrill-Palmer Quarterly, 1961, 7, 147-169.

Torrance, E. P. Rewarding creative behavior. Englewood Cliffs, New Jersey: Prentice-Hall, 1965.

Upton, A., and Sampson, A. V. Creative analysis. Whittier, California: Whittier College Press, 1961.

Measuring Motivation in Culturally Disadvantaged School Children

Richard deCharms
Washington University
St. Louis, Missouri

Virginia Carpenter
Lindenwood College
St. Charles, Missouri

A COLLEAGUE of mine recently spent a semester in seventh grade (Smith and Geoffrey, 1965). In a school serving a culturally disadvantaged neighborhood, he tried to attend class every day and played the role of a nonparticipant observer. His account of this experience is very discouraging to a person who is primarily interested in human motivation. Take, for instance, one simple incident:

> Harry had been over at the confectioner" across the street, and had come out with the rest of the boys on Tuesday, when a fight began on the playground. There had been considerable excitement and Mr. Inman (the principal) was trying to calm everyone down. Harry had left his pen over in the confectionery and wanted to go b- - across the street to get it. Mr. Inman .ad told him to wait for a few minutes. A little while later, he asked the question again and was told to wait. Then, or on the third time he had asked and was turned down and told not to bother Mr. Inman, Harry walked off and made a muttered and indecent comment. This was the end of it for him.

This commonplace incident in itself is minor and doesn't carry the impact of the whole report where such incidents cumulate. It can be taken, however, as illustrative. The utter futility of the situation is striking. This and other incidents make it clear that the children feel powerless to control their own destinies. As a result of their futile attempts to assert their initiative in inappropriate ways, they create an environment where even the teachers and administrators are forced to act in ways that are not of their choosing. The result is a constant struggle between pupils and the school staff to obtain the upper hand. Teachers search for ways to enhance their power to control the behavior of pupils, and the children use every means at their command to subvert the authority of the school, the last resort being hostility through passive dependency. Both parties to the situation are ultimately pawns in the hands of the powers beyond their control.

Two more or less disconnected things stand out for me in approaching the problem of motivation in school children. First, the tangled web of influences will not yield to any miraculous new insights or techniques. The Gordian knot is real and will only yield very slowly to modest and patient efforts to unravel one or two threads at a time. In this age of crash programs, imminent breakthroughs, etc., motivation is an important concept to consider, but we must not expect it to solve all problems.

The second thing that strikes me is that one of the loose strands that may have potential for unraveling both practical and theoretical aspects of the knot is the notion of powerlessness, which is, after all, a motivational concept.

What I have to report then, is no breakthrough, but rather a modest beginning that attempts to discover whether certain aspects of motivation can be measured in a population of culturally disadvantaged children, and whether some of the motivational concepts akin to powerlessness can be used.

Actually we will discuss three different but related things. First, I will present some notions about powerlessness and what I will call the "Origin-Pawn" variable and discuss very briefly some relevant laboratory studies that we have completed. Second, I will shift to a discussion of achievement motivation and some of the conceptual and scaling difficulties that it presents. This is necessary in order to introduce a third section dealing with data on the achievement motive in culturally disadvantaged children. These three sections may seem a bit disjointed. What holds them together is the relationship between the concepts to be presented in section one and the achievement motive. At the present stage of the research there is little to report on these concepts except laboratory studies, but we do have evidence on achievement motivation from culturally disadvantaged children. As a result, I have more or less arbitrarily put the two together with a discussion of some theoretical problems with the measurement of achievement motivation.

POWERLESSNESS AND THE ORIGIN-PAWN VARIABLE

In the anecdote related above, Harry was frustrated by his powerlessness to influence an important aspect of his life vis-à-vis a school official. He was a pawn to forces beyond his control. In much of his experience in life, and especially in school, he is, in fact, and is expected to be, a pawn. Learning seems to demand that he react to outside influences in the school, and one way to insure that he will react in the proper manner is to reduce his freedom to a minimum, to force him to study, to press him into the mold expected by the school. Techniques designed to do this demand that the teacher and the school have most of the power and that power be taken away from the student.

Motivationally there is something wrong with this picture. The student, caught up in a power field, will do everything he can do to break out of it, and when he leaves school, he will be out of it. Is the goal of teaching to produce certain responses in the school that are almost certain to be a function of the specific power situation and, therefore, not transferred to non-school situations? Or is there a broader goal?

We would like to distinguish between the driven or power-induced school behavior, and behavior motivated from within the student that will carry over to extra-school behavior. The belief that the latter is possible reduces to a naive assumption that we all make in our everyday interaction. The assumption is that each individual has some autonomy and exercises it by originating his own behavior. Let me use the one word "Origin" to connote a person who is perceived to be originating his own behavior. A person is not always an "Origin." Often we perceive his behavior as stemming from a force field so that he is coerced to behave in a certain way by an external source. In such cases, I shall refer to him as a "Pawn" or as acting like a "Pawn." It should be clear that pawn-type behavior is induced by lack of power on the part of the pawn. Origin type behavior implies a feeling of power, although it is not entirely complementary in this respect to pawn-type behavior. Pawn behavior implies powerlessness, but origin behavior implies freedom of choice, not necessarily the ability to wield power over others.

Having sketched the "Origin-Pawn" dimension, let us note that it is essentially a phenomenological given in the perception of behavior. As such, it may be a myth or a mental way station, to use Skinner's railroading metaphor. To extend the metaphor, we might say that the main line of behaviorism runs nowhere near this way station, and efficient operation of the line would suggest that the station should be dropped. In fact, scientific psychology has been at pains to discredit the assumption that a human being can be the ultimate origin of some of his behavior.

Behaviorism, suffering under the strong influence of operationism and physicalism, has done its best to reduce the concept of *motivation* in psychology to something of the order of "the study of the causes of behavior." The trouble with this approach is that trying to elucidate the concept of motivation by means of the concept of cause is like trying to understand the behavior of a human being using concepts found to be inadequate to explain the behavior of billiard balls. The precept that all scientific statements must be based ultimately on statements about physical operations has great merit, but the hidden implication that explanations of physical motion (including behavior) are reducible to the laws of mechanics has seriously retarded the development of theories of human motivation.

Stripped of its hidden assumptions, the neo-behavioristic approach to motivation as the cause of behavior is unacceptable to anyone and has never been seriously held by anyone. It assumes that the human being is like a billiard ball and his behavior, like that of the ball, is entirely the result of external sources forcing him to behave. Another object hits him and steers him in a certain direction – an explanation that would account for certain trivial behaviors on the football field perhaps – or an external noxious stimulus repels him like one negative charge repels another, or an internal condition (e.g., drive-stimulus) produced by external circumstances (e.g., food deprivation) drives him in a certain direction.

Confronted with the phenomena of motivated behavior which in this framework appears akin to a sudden inexplicable movement of a billiard ball resting alone on a perfectly level table, psychology has ingeniously devised many schemes to explain this untoward behavior in terms of some mechanisms involving a causal chain of events and ultimately reducing to some undetected source of physical energy closely connected with, or actually residing in, an impelling stimulus (e.g., the cue ball). The cue ball, or impelling stimulus, has been variously described as a need or drive (Hull, 1943), a drive-stimulus (later Hull and Miller), incremental stimulation (Mowrer, 1960), or an affective stimulus (Young, 1961; McClelland, 1951).

In the face of all this theorizing based on what Gilbert Ryle (1949) would call a para-mechanical myth, it seems to me that Skinner has done the most consistent thing for a strictly empirically based "science of behavior," – namely to shun the

notions of causality and motivation altogether. Although I suspect that he may have led some to hide their motivational and causal notions under the concept of reinforcement, his strict adherence to the definition of a reinforcer as something that increases the probability of a response does, in fact, rid the science of behavior of all of these irritating problems.

By now, I hope to have laid the groundwork so that two conceptions of man as a motivated being are clear. The first is the one we use in our everyday dealings with human beings: namely that the cause or reason for a person's behavior is sought first within him. More often than not, he is the origin of his behavior and is to be held responsible for it. The second is the explanation from scientific psychology that seeks a physical source external to the person, or at least a stimulus within him, that drives him. This latter notion has never been satisfactory in accounting for the phenomena of human motivation and has, through its emphasis on demonstrating a physical basis for all behavior, led to a psychology of learning and motivation that conceives of the animal and the human being as a pawn to be manipulated by reinforcement schedules.

Accepting at face value the psychological fact that people perceive themselves and others as origins in most instances is, for me, the starting point of a truly psychological approach to motivation, and demands a level of analysis quite different from the causal chain analysis that attempts to reduce the explanation to physical events. I have tried, therefore, to build on the assumption that the origin-pawn dimension makes a difference that is measurable in the behavior of a subject.

The origin-pawn dimension has its roots in the concept of locus of causality used by Heider (1958). His emphasis is on the perception of causality, and the result is that the concept is applied in the area of person perception. In perceiving another person's behavior, we do attempt to assess whether the locus of causality is internal or external to the person, and this distinction affects our perception of the person, the characteristics that we attribute to him, and the way we behave toward him.

Several person perception studies have demonstrated the significance of attribution of locus of causality (Thibaut and Riecken, 1955; Pepitone, 1958). In our own research, we have demonstrated (deCharms, Carpenter, and Kuperman, 1965) that subjects clearly do use the origin-pawn variable in their perceptions of others. We found that the perception of a person as an origin or as a pawn is a function of several variables, such as whether some agent is attempting to influence the person's behavior or not. Origin or Pawn perceptions are a function of characteristics of the influence agent, such as whether it is a large group, a small group, or an individual, and whether it is liked or disliked by the person. Most interesting, however, was the finding that a personality characteristic of the subject strongly affected his perception of another person along the origin-pawn dimension. Subjects were measured on a scale developed by Rotter, Seeman, and Liverant (1962) that attempts to measure the general feeling of powerlessness or efficacy of the person. Since Rotter uses reinforcement terminology, he calls the scale a measure of control of reinforcements and names it the I-E scale standing for internal or external control. There is a subtle difference between this concept of control of reinforcements and our O-P dimension, but for now the similarities may be noted. Our data showed that subjects who reported on the Rotter I-E scale that they felt in control of their sources of reinforcement perceived other people more as Origins, whereas subjects at the other end of the I-E scale perceived other people more as Pawns.

Since our primary interest is motivation, we next asked the question: "What effect does feeling like an Origin, as opposed to feeling like a Pawn, have on the person's own behavior?" With this question we moved from the perception of others to the perception of self as an origin or a pawn. Two laboratory studies (Kuperman, 1966; deCharms, Dougherty, and Wurtz, 1965) were devised that attempted experimentally to induce the person to feel like a pawn in one situation and like an origin in another, and to measure the effects on their behavior. The results confirmed the influence of the origin-pawn variable as manipulated in the experiment. Subjects building a tinker toy model under origin conditions that attempted to induce a free atmosphere, as compared to building a similar model under pawn conditions a) liked the origin model best, b) became more involved, c) chose to continue the origin model when interrupted, d) completed the origin model more elegantly, and e) recalled the nonsense name given the origin model more frequently a month after the experiment. Apparently, the origin-pawn dimension can be manipulated in a way that has strong effects on behavior. The Rotter I-E scale was administered to these subjects also, but was not related to any of the dependent variables, either directly or in interaction with the origin-pawn manipulation. Evidently, this questionnaire measure relates to other questionnaire responses as in the previous study of person perception, but doesn't predict behavior related to a task situation, at least for the male subjects used here.

ACHIEVEMENT MOTIVATION

So far our work on the Origin-Pawn variable has not ventured very far out of the laboratory, and we have very little empirical evidence about the variable in culturally disadvantaged school children. What we do have, however, is preliminary evidence about the measure of achievement orientations in this population, and some evidence that one of the aspects of achievement orientation involves acting like an origin. Let us be clear that I am not equating acting like an origin with having high n Achievement. What I am doing is looking for a relationship between the two concepts, and since I am just launching a project to attempt to measure the O-P variable and n Achievement in schools, and already have data on n Achievement, I will report the data in hand on n Achievement and speculate about the O-P variable. The connection is closer than may be apparent.

McClelland (1961) has spelled out several aspects of what he called the achievement syndrome, i. e., characteristics of people with high achievement motivation. One of the most important aspects is that

the achievement motivated person likes to take personal responsibility for his actions. He likes to be in a position to act like an origin and avoids situations where he is a pawn.

Let us shift our emphasis, then, to the measurement of n Achievement in culturally disadvantaged children. First, let us be very clear about one thing. The word "achievement" is a loaded term, especially in educational circles. A culture such as ours, that has been nurtured on Horatio Alger stories and McGuffey readers, tends to perceive achievement motivation as an unmitigated virtue if not a blessing. If we pull back from this conception and ask what aspect of achievement motivation we can actually measure, we may be disappointed. The measure to be used was developed about fifteen years ago by McClelland and his associates at Wesleyan University, and the subsequent years of research have told us much about the characteristics of people who score high on the measure. The most appropriate single word to describe the person with strong achievement motivation as measured by this technique is that he is an "entrepreneur," typically following a business career. Topcreative scientists are not particularly high on n Achievement, nor are academicians. If you want to succeed in business, however, you'd better have high n Achievement.

The achievement syndrome has three major aspects: 1) personal responsibility, 2) risk-taking strategies, and 3) the use of feedback. The entrepreneurial spirit of achievement emphasizes self-reliance, the taking of calculated risks, careful planning and checking of progress with constant emphasis on the skillful use of one's abilities. These, then, are the dependent variables that are related to the measure of achievement motivation.

What is the measure itself? Actually, even after years of research the original form of the measure has not been improved very much and is essentially a very crude measure. Originally intended as a measure of individual differences, it yields at best a very rough ordinal scale that is most useful when collapsed into a high-low dichotomy, or at most a high-medium-low trichotomy.

Before discussing the weaknesses of the measure, let us sketch the procedure for obtaining n Achievement scores. The technique involves collecting thought samples from subjects under standardized conditions by showing a group of subjects several pictures, one by one, for a short period of time and asking them to write a creative story about each in turn. The stories are content analyzed by trained coders for instances of Achievement Imagery. Stories in which one of the characters is concerned over obtaining success in competition with a recognized standard of excellence are said to contain achievement imagery. The construct validity of the measure is attested by years of research although test-re-test reliabilities are often lower than validity coefficients (for reasons that we cannot take up here).

Of basic concern to us is the question: Does this measure have any utility for the measurement of motivation in educational settings and specifically in culturally disadvantaged areas? Most of the early research results were disappointing in this regard. Apparently, n Achievement scores are not very highly correlated with school grades or standardized achievement test scores in high-school and college students. As a result, educational interest in the measure lagged after an early flirtation, but it has recently been revived by the evidence that it is possible to change the motive by training. Before we get dazzled by the thought of changing motives in education, we should try to understand the relationship between n Achievement and school behavior. I think there are good reasons to believe that the measure has been misapplied but that it, or a derivative of it measuring the O-P variable, may be very valuable for education.

This misapplication of the n Achievement measure stems from the two most basic levels possible: a lack of understanding of the theoretical basis for the measure, and lack of clarity about its scaling properties.

The problem of understanding the theoretical basis for the n Achievement measure is formidable. Exactly why there should be a relationship between the content of stories written by a subject in a one-half hour period and many other manifestations of his behavior, particularly school behavior, is still a problem shrouded in mystery. I will spare you a discussion of several explanatory attempts that have been made and present one based on the notion that the measure gives a sample of thoughts of the subject.

Let us be very simple-minded and conceive of thoughts as a type of response, arbitrarily remove them from the ghostlike realm of the mind, and relegate them to the more prosaic realm of the physical world of behavioral events. After all, the n Achievement measure is actually based on observable responses, not on pure thoughts. From this point of view, thought samples are a measure of one domain of responses that we wish to correlate with other domains. Notice in passing that we have made no assumptions about a causal chain, that thoughts are antecedent to or "cause" behavior. We have a habit of thinking that thoughts precede actions but there is very little evidence to substantiate it, and in the realm of achievement motivation, there is some evidence that the behavior may precede the thoughts (McClelland, 1966).

The definition of what constitutes an achievement thought is critical in predicting what behaviors will be associated with such thoughts. The scoring manual (Atkinson, 1958) identifies achievement thoughts by the term Achievement Imagery and says they can be seen in a story when one of the characters shows concern over competition with a standard of excellence. The character can show competition when any concrete goal is at stake, such as winning a race, getting a good gra e, becoming a success in a career, building a better mousetrap, and so on. The non-specificity of the goal of achievement behavior leads us to the scaling problem mentioned earlier to which we shall return, but let us pursue a little further the theoretical basis for the measure.

Thought-sampling is based on an assumed correlation between thoughts and more concrete behavior.

The advantage of using a sample of thoughts rather than some other type of behavior is this: of all behaviors, thoughts are least subject to inhibition. If we seek a measure of what a person wants to do, we need to rid it of restrictions as to what he can do. For this and other reasons, thought-samples stand a chance of giving us a better picture of a person's generalized motives.

In attempting to measure what a person wants to do, we must follow quite a different road from that taken in measuring what a person is able to do. Put in another way, the measurement of motivation is quite different from the measurement of intelligence, ability, or even attitudes. We want to know what a person will do, not what he can do, although the latter may form a limit in any particular circumstance. In measuring motivation, we sample his thoughts under neutral conditions: not when he is pressured and a pawn, but when he can originate some of his own behavior. By definition, in such a situation, we cannot have control of the kinds of responses observed. If we tighten control, we lose the ability to measure what he will do on his own and end up with a measure of what he can do when we pressure him.

This paradoxical measurement situation is more or less unique and doesn't fit very nicely into some of the most basic conceptions of scaling. To make this a little clearer, think of Skinner's distinction between operants and respondents. Roughly, a respondent is a response that we can elicit directly by manipulation of a specific stimulus. We more or less directly produce a response in the organism by external manipulation. Salivary, pupillary, and eyelid responses are examples. An operant is quite different. Although Skinner assumes that this type of response is probably elicited by some stimulus, we don't know and may never be able to know what it is, so we cannot control the response through the eliciting stimulus. Ignoring control by reinforcement for the time being, we measure an operant by observing its spontaneous emission under conditions of relative freedom.

Let us now conceive of thought sampling as measuring the rates of spontaneously emitted thoughts, i. e., as the measurement of operants. Here we launch into new territory. While Skinner is primarily concerned with shaping a specific operant and converting uncontrolled behavior into behavior that is under the control of the experimenter, we are concerned with measuring a generalized class of operants in the realm of thoughts and relating this measure to other more concrete behavior patterns.

This leap raises many problems, but the major methodological problem is that our measure of a generalized class of spontaneously emitted thoughts cannot be assumed to result in a conventional scale characterized by transitivity along an intensity dimension. Let me try to illustrate this: Pictures used to elicit thought samples from which n Achievement scores are derived show a variety of characters, for instance, a high-school student, two inventors, some businessmen, etc. In what sense can a person (A) who writes achievement stories about inventors, doctors, and businessmen and attains a high score be said to be higher on n Achievement than another person (B) who writes an achievement story about a high-school boy but not about inventors, doctors, and businessmen and receives a relatively low score? If we observe many types of behavior in many situations, we may assume that person A will show more general achievement behavior than person B. But if we measure only school behavior, the reverse may be true. In an unpublished study, we have demonstrated that a six picture measure of n Achievement derived from a standard set does not meet the criterion set by the Guttman technique for unidimensional scalability (Stouffer, Guttman, Suchman, Lazarfeld, Starr, and Clausen, 1947).

Does this mean that the measure is simply unreliable? I think not, although it does mean that it is not a unidimensional measure of the intensity of a motive to do a very specific thing and, therefore, should not be expected to relate to or predict rather specific behavior, such as that expected on a standardized achievement test taken under conditions that produce respondent, or at least discriminated operant, rather than spontaneously emitted operant type behavior. It would be nice to be able to measure motivation along an intensity scale, and most of motivation theory influences us to think of motivation in terms of intensity concepts such as need and drive. If we think back for a moment to the discussion of the Origin-Pawn dimension, however, we can see that a generalized measure of what a person will do when not forced may come closer to our concept of motivation than an intensity measure of what a person can do when pressured.

In the book presenting the original development of the n Achievement measure, McClelland, Atkinson, Clark, and Lowell (1953) briefly mentioned that the technique measured the extensity of Achievement thoughts in a broad realm rather than the intensity of such thoughts. This mention has apparently been lost in the flood of research, most of which was not careful to note the the distinction.

As suggested above, a careful analysis of the basic process involved in the measurement technique gives some insight into the low correlations between n Achievement and academic performance. It gives us more. I think it gives us the basis for suggesting that n Achievement scores will be related to much more general response patterns. Since the measure taps thoughts about achievement in several areas, if a person who has many such thoughts is observed over a period of time in which he may be confronted with some of these situations, we should be able to predict that he will react in achievement ways more often than not. In view of the analysis, it may be expected that the measure of n Achievement will be more valuable in relation to cultural orientations (McClelland, 1961) and long term career patterns of individuals, than in predicting specific responses such as school achievement or exam-taking behavior.

One example may be cited. McClelland (1965) reports that of the original sample of subjects tested at Wesleyan in the validation procedures for n Achievement fifteen years ago, large numbers have gravitated toward entrepreneurial business occupations subsequent to graduating from college. A check of the alumni directory indicated that 83 percent of the men who could be classified as entrepreneurs now, had high n Achievement scores as sophomores in

college. Only 21 percent of those classified as non-entrepreneurs had high n Achievement. These results indicate that n Achievement predicts career patterns over a crucial period of life when career choices are being made.

When it comes to measuring motives in culturally disadvantaged children, we have a lot to learn. The data that I have to report results from a study that was initially intended to answer some of the most rudimentary practical questions such as "Can thought samples be collected from this population, i. e., can they produce scorable stories under standardized conditions?" If it turns out that the thought samples are scorable, we can ask whether the n Achievement score is valid for this population. Is the score, for instance, related to risk-taking strategies in this population as it is in middle-class high-school and college students? Since we were interested in beginning to probe the Origin-Pawn dimension in this population, we sought an existant measure that would give an estimate of the children's relative feelings of powerlessness. In short, we tested the feasibility of collecting thought-samples and found fifth, sixth, and seventh grade disadvantaged children quite capable of producing scorable stories. We devised a risk-taking situation that had the bonus of providing us with a measure of school skills, and we used a measure of powerlessness.

The subjects of the study were ninety-four Negro children in two fifth and seventh grade classrooms in one large city (sixty girls and thirty-four boys), and 120 Negro children from several sixth grade classrooms in another large city. Both samples were taken from school areas known to be the most culturally disadvantaged in each city. However, since we were primarily interested in testing the feasibility and validity of the motive measure, we did not sample meticulously, nor did we carefully assess the comparability of the areas. For this reason, comparisons between the two samples are dangerous.

THE MEASUREMENT OF ACHIEVEMENT MOTIVATION

After introductory remarks, the children were handed a six-page form with a different cue typed at the top of each page, and four standard questions equally spaced on each page to help them complete their stories. The verbal cue was read aloud (e. g., "Two men are working at a machine."), and the children were given four minutes to write their stories. In all, six stories were required.

These stories were content analyzed by the procedures described by McClelland, et al. (1953) by two independent scorers. Interscorer reliability was approximately 90 percent agreement between scorers.

THE MEASURE OF POWERLESSNESS

The children's form of the measure of powerlessness (Bialer, 1961) was administered immediately after the measure of achievement motivation. Twenty-three statements are presented in simple language and the child circles either "Yes" or "No" indicating whether he agrees or disagrees with the statement. For example, one item was, "Does it ever help any to think what you will be when you grow up?"

RISK-TAKING BEHAVIOR

In a game situation in which the person can choose to shoot for a fixed target from varying distances of increasing difficulty with increasing reward attendant on success, a group of high-achievement persons will tend to take more shots in the middle range, avoiding both easy and extremely hard shots. A similar group of low-achievement persons will take more shots in the easy and extremely hard positions (cf., Atkinson, Bastian, Earl, and Litwin, 1960). The range of risk which may be considered a moderate risk is dependent upon the skill of the individual and is evidently chosen in terms of where he thinks he has a moderate chance of success (cf., deCharms and Davé, 1965).

We rejected the simple game situation most often used. Instead, it was decided to attempt to develop a new situation completely analogous to the game utilized in our earlier work but involving the typical school behavior of children, i. e. spelling and arithmetic skills.

SPELLING AND ARITHMETIC SKILLS

Each of these measures was comprised of sixty items divided into six levels of difficulty of ten items each. Spelling words were taken from standard lists calibrated for difficulty, and arithmetic problems were taken from a series of arithmetic textbooks used in the schools from early elementary-grade level to junior-high-school level. Extensive pretesting was used to assess empirically the validity of the various levels, and to assure a high degree of success for all children in the easiest level, and increasing difficulty up to a level where little success could be expected from any child.

These measures were administered by standard classroom procedures. The items were arranged in blocks of six, each block progressing from the easiest to hardest level of difficulty in order, though this was not made obvious to the children. It was stressed that this was not a test in the sense that the children would receive grades, but that they themselves would learn from this about their own skill.

After all children had attempted all sixty items, the various levels of difficulty were explained to them and their own ability at each level was impressed upon them by pointing out how many they had completed correctly at each level.

SPELLING AND ARITHMETIC RISK-TAKING TASK

After the group session, the experimenter met with each child individually. She made sure that he understood the various levels of difficulty in the sixty items he had attempted, and how many he had completed correctly at each level.

The child was told that he would now take a test in which he could choose each time the level of difficulty of the item which he would attempt. It was made clear that the test items were divided into levels comparable to those of the previous skill task.

Before he was presented with each item, he was to choose the level of difficulty which he would like to attempt on that trial. The object was to attain "points," the significance of which was left unexplained, as in a game. The points that could be won by correctly completing the item increased with the level of difficulty of the item chosen. These points were established for each child separately, according to the empirically established level of difficulty for him. Thus, a skillful child who chose level five (a difficult level), and who had done four out of ten items correctly at that level in the skill-task (empirical probability of success equals .40), stood to gain thirty-eight points if correct. A less skillful child choosing the same level five may have done only one out of ten items correctly at that level in the skill-task (empirical probability of success equals .10). Such a child could win 150 points if correct at level five. (This procedure, developed by deCharms and Davé (1965), equates subjects as to skill and emphasizes choices in terms of the child's own chances of success.)

Each time the child chose a level for the next item (of ten spelling and ten arithmetic items), he was asked to consult his own paper from the skill-task and report how many he could expect to get right out of ten (his probability of success at that level). He was then asked to consult the following table and report how many points he could win. The point scale was constructed so that the number correct multiplied by the number of points always approximated a constant (150) as closely as possible without using decimals. The expected value of ten trials at any level, therefore, equals 150.

Number you got right out of ten on skill-task	10	9	8	7	6	5	4	3	2	1	0
Points you get if right	15	17	19	22	25	30	38	50	75	150	300

By such means the experimenter made every effort to insure that the child thought of his choices in terms of his own performance on the skill-task and in terms of the possible points which were assigned in accordance with his own demonstrated skill.

Each child made ten choices and attempted ten spelling words, recording his accrued points after each trial. Following this, each child did the same with arithmetic items.

RESULTS

Table 1 presents the mean skill scores on arithmetic and spelling tasks by sex and level of achievement motivation. You will note that the means for the high n Achievement subjects are higher in all cases than the means for low n Achievement subjects. The differences tested by analysis of variance are significant in every case except for arithmetic for girls. There is clearly a relationship overall between n Achievement level and school skills as measured by our test. The relationship is stronger for boys than for girls, but all differences are small, a fact that is attributable to the low variances of the scores. The 60-item task was designed so that the majority of the subjects would success on almost all of the items at the easiest level and fail on almost all at the most difficult level. As a result, the range of probable scores is much smaller than the total range of sixty, and the variances are quite small. In addition, these results come from a larger analysis where variance attributable to other factors, such as grade level, was extracted.

Table 2 presents data from the same dependent variables, spelling and arithmetic skills, but this time showing the relationship with the I-E scale. The low I-E subjects feel relatively powerless. The relationship is significant in both cases of arithmetic and spelling skills for girls but not for boys. This sex difference is more or less the opposite of what was found with n Achievement. These results seem to indicate that the thought sample measure is more powerful in predicting school skills for boys, whereas the questionnaire measure is more powerful for girls. The data by themselves are not very impressive in this respect, and might seem merely a function of the specific tests except for two things. First, we have argued that thought sampling results in a different kind of score from that derived from a questionnaire, and the difference is interesting theoretically. Second, a recent review of the literature of n Achievement indicates a similar trend toward an interaction between sex and type of testing situation. Klinger (1966) reviewed some forty-four studies that investigated the relationship between measures of achievement motivation and task performance. Some of the studies used the thought-sample technique which Klinger calls the TAT, and some of them used a measure called the Iowa Picture Interpretation Test, or the IPIT. This test uses pictures but supplies several plots for the subject to choose from in a multiple choice form. Klinger says: "The studies also suggest an interaction of Instrument × Sex of Subject, in that of the studies that used the TAT, a higher proportion of those with male subjects reported significance than those with females, while the reverse was true with the IPIT" (Klinger, 1966, pp. 297-298). Klinger makes no attempt to explain this interaction.

In view of our analysis of thought sampling as a measure of a more or less "free" operant where the subject emits responses, whereas a questionnaire or multiple choice measure supplies ready-made responses for the subject to choose from, it seems reasonable to hypothesize that the free operant measure may be more appropriate for male subjects and a respondent (or discriminated operant) measure more appropriate for females. I cannot pretend to explain in detail why this should be so, but it does seem plausible, especially in the achievement area, and in our culture, that men are expected to produce achievement behavior spontaneously, whereas women may have to respond appropriately with achievement behavior when pressed but are not necessarily expected to initiate it on their own. We hope to investigate this possibility in the near future.

The results from the risk-taking data ar[illegible] presented in Tables 3 and 4, and may be seen ir[illegible] ures 1 and 2. The means in the tables need a li[illegible] e explanation. The hypothesis derived from earlier studies predicts that subjects with high achievement motivation will more often choose spelling words or

TABLE 1

MEAN SPELLING AND ARITHMETIC SCORES (OF A POSSIBLE 60) BY SEX AND LEVEL OF n ACHIEVEMENT

		Boys			Girls		
		Math**	Spelling**	Total	Math	Spelling*	Total
High	N	46	46		60	60	
n Achievement	$\bar{x}$	21.7	28.2	24.9	22.5	31.0	26.8
Low	N	51	51		63	63	
n Achievement	$\bar{x}$	18.5	21.7	20.1	19.6	26.5	23.1
Total		20.0	24.8	22.4	21.0	28.7	24.9

* $p < .05$
** $p < .01$

TABLE 2

MEAN SPELLING AND ARITHMETIC SCORES (OF A POSSIBLE 60) BY SEX AND I-E LEVEL (POWERLESSNESS)

		Boys			Girls		
		Math	Spelling	Total	Math*	Spelling*	Total
High	N	46	46		69	69	
I-E	$\bar{x}$	20.9	25.2	23.1	22.6	29.8	26.2
Low	N	51	51		54	54	
I-E	$\bar{x}$	19.3	24.4	21.9	19.0	27.2	23.1
Total		20.0	24.8	22.4	21.0	28.7	24.9

* $p < .05$

arithmetic problems in the middle ranges of probability levels from say 30 percent to 70 percent probability, while subjects with low n Achievement will not show this tendency to as great an extent and may, in fact, avoid the middle ranges and choose more extremes either around 90 percent or 10 percent.

In order to extract a measure from the data to test this hypothesis statistically, we used a procedure developed by deCharms and Davé (1965). The average deviation of the ten probability choices for each subject is measured around the overall mean of all choices for all subjects. A large average deviation indicates that the individual subject's choices range farther out in either direction on the scale; a small average deviation indicates that the subject's choices are concentrated in the middle range. The hypothesis then predicts that subjects with high n Achievement will have smaller average deviation scores than subjects with low n Achievement. Tables 3 and 4 present the means of the average deviation scores from each group. We shall call this score the risk score.

Table 3 gives strong support for the hypothesized relationship between moderate risk-taking and high n Achievement under all conditions, i.e., for both boys and girls and when the task involves either spelling or arithmetic problems. Table 4 indicates that the I-E scale is not related to risk-taking under any conditions.

A look at Figures 1 and 2 may make the results with n Achievement clearer. Figure 1 combines

TABLE 3

MEAN RISK-TAKING SCORES BY SEX AND LEVEL OF n ACHIEVEMENT
(LOW MEAN INDICATES MODERATE RISK)

		Boys			Girls		
		Math*	Spelling***	Total	Math*	Spelling***	Total
High n Achievement	N	46	46		60	60	
	x̄	19.2	24.2	21.7	22.0	23.3	22.6
Low n Achievement	N	51	51		63	63	
	x̄	23.8	29.1	26.4	24.4	27.9	26.1
Total		21.6	26.8	24.2	23.2	25.6	24.4

* $p. < .05$
***$p. < .005$

TABLE 4

MEAN RISK-TAKING SCORES BY SEX AND LEVEL I-E (POWERLESSNESS)
(LOW MEAN INDICATES MODERATE RISK) *

		Boys			Girls		
		Math	Spelling	Total	Math	Spelling	Total
High I-E	N	46	46		69	69	
	x̄	21.7	27.5	24.6	22.7	25.3	24.0
Low I-E	N	51	51		54	54	
	x̄	21.5	26.2	24.4	23.8	26.0	24.9
Total		21.6	26.8	24.2	23.2	25.6	24.4

*No significant differences

boys and girls and shows a high peak in the mid ranges of probabilities of success for the high n Achievement subjects on arithmetic problems. The curve for the low n Achievement subjects shows an interesting propensity to take extremely speculative risks. For the low n Achievement subjects 58 percent of the choices are in the range of 0 to 40 percent probability of success.

Figure 2 presents the results for spelling. Here the high n Achievement curve shows a hump in the middle, and the low n Achievement curve actually shows a trough in the middle ranges and is higher at both extremes as predicted by the hypothesis.

These results with arithmetic and spelling tasks show a clearer differentiation in risk-taking strategies between high and low n Achievement subjects than most previous studies in this area, and suggest the possibility that risk-taking on a serious task such as spelling or arithmetic is more clearly a function of n Achievement than on the game-like tasks used in other studies. It is also possible that the effect of achievement motivation may be more clearly evident in the culturally disadvantaged environment than in the highly competitive middle class culture where the range of n Achievement scores may be skewed toward high n Achievement.

DISCUSSION

All of the results reported from this study were derived from two different samples that were combined for simplicity of presentation. When the samples are investigated separately, the findings reported replicate in almost every case.

Figure 1. Percentage of tasks chosen at each level of probability of success on the arithmetic risk-taking task by subjects above and below the median *n* Achievement score.

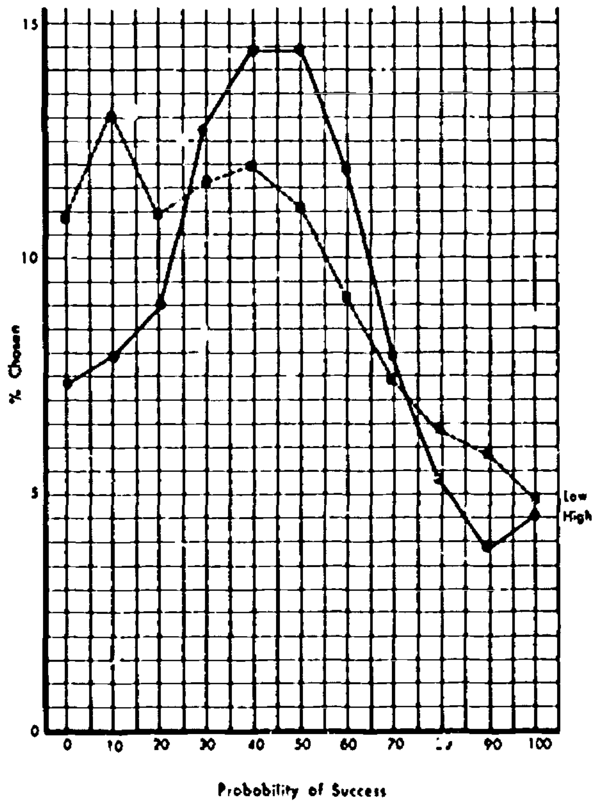
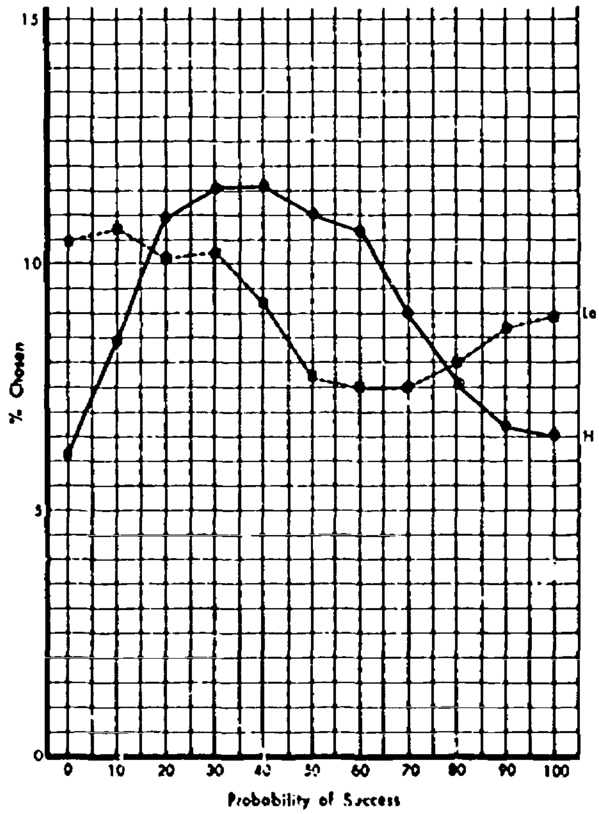


Figure 2. Percentage of tasks chosen at each level of probability of success on the spelling risk-taking task by subjects above and below the median *n* Achievement score.



Replication of the data on spelling and arithmetic skills is especially important since the tests from which these scales derived were intended originally merely to give preliminary data for the risk-taking task. The fact that the relationships with *n* Achievement are significant at all is surprising and even embarrassing since the major thrust of our theoretical argument predicts that the *n* Achievement measure should be more highly related to measures of change over time than to single test measures such as these arithmetic and spelling skill tasks. Actually, the results, though significant, are not indicative of high correlations, but of relationships quite commensurate with others reported between *n* Achievement and academic achievement. Although we have argued that *n* Achievement is primarily a measure of the extensity of thoughts about achievement and should, therefore, be related to long-term trends in behavior, our data remind us that a measure of extensity is not completely independent of intensity. To use the extensity measure exclusively as a measure of intensity is to misapply it; but the extensity of thoughts about achievement may give some indication of the intensity of achievement motivation in a specific area.

Probably the most interesting result of these preliminary forays into the study of motivation in culturally disadvantaged children is the evidence that the thought sample measures are more effective with boys than girls, and with the questionnaire measure the reverse is true. If this result stands up on replication, it may lead to further insights about thought sampling.

It should be clear by now that the I-E scale used here was not designed to measure the origin-pawn dimension. It is our hope that a measure of this dimension can be devised based on thought sampling. To be valid, the measure must allow the subject to emit his own behavior. Choice of reactions selected and presented by the experimenter as on a questionnaire limits the subject. We are at present beginning to develop a measure based on thought samples, and preliminary evidence indicates that it is feasible and may be related to the measures of school skills discussed here and even to long-term indications that a person prefers to take personal responsibility for his actions.

We can only end this paper by noting that it is only a beginning. The research reported here constitutes the basis for a project that is just getting under way, the aims of which are to relate *n* Achievement to change in academic skills in a longitudinal study, to develop the concept and measure

the origin-pawn dimension, and ultimately to develop training techniques for teachers and pupils similar to those used in achievement motivation training courses with businessmen (McClelland, 1965).

REFERENCES

Atkinson, J. W. (Editor) Motives in fantasy, action, and society. Princeton, New Jersey: Van Nostrand, 1958.

Atkinson, J. W., Bastian, J. R., Earl, R.W., and Litwin, G. H. The achievement motive, goal setting, and probability preferences. Journal of Abnormal and Social Psychology, 1960, 60, 27-36.

Bialer, I. Conceptualization of success and failure in mentally retarded and normal children. Journal of Personality, 1961, 29, 303-320.

deCharms, R., Carpenter, Virginia, and Kuperman, A. The "origin-pawn" variable in person perception. Sociometry, 1965, 28, 241-258.

deCharms, R., and Davé, P. N. Hope of success, fear of failure, subjective probability, and risk-taking behavior. Journal of Personality and Social Psychology, 1965, 1, 558-568.

deCharms, R., Dougherty, K., and Wurtz, S. The origin-pawn variable. Unpublished manuscript, Washington University, 1965.

Heider, F. The psychology of interpersonal relations. New York: Wiley, 1958.

Hull, C. L. Principles of behavior: An introduction to behavior theory. New York: Appleton-Century, 1943.

Klinger, E. Fantasy need achievement as a motivational construct. Psychological Bulletin, 1966, 66, 291-308.

Kuperman, A. Relations between differential constraints, affect, and the "origin-pawn" variable. Unpublished doctoral dissertation, Washington University, 1966.

McClelland, D. C. Personality. New York: Sloane, 1951.

McClelland, D. C. The achieving society. Princeton, New Jersey: Van Nostrand, 1961.

McClelland, D. C. Toward a theory of motive acquisition. American Psychologist, 1965, 20, 321-333.

McClelland, D. C., Atkinson, J. W., Clark, R. A., and Lowell, E. L. The achievement motive. New York: Appleton-Century-Crofts, 1953.

Mowrer, O. H. Learning theory and behavior. New York: Wiley, 1960.

Pepitone, A. Attributions of causality, social attitudes, and cognitive matching processes. In R. Tagiuri and L. Petrullo (Editors), Person perception and interpersonal behavior. Stanford: Stanford University Press, 1958.

Rotter, J. B., Seeman, M., and Liverant, S. Internal versus external control of reinforcements: A major variable in behavior theory. In N. F. Washburne (Editor), Decisions, values and groups. Vol. 2. New York: Macmillan, 1962.

Ryle, G. The concept of mind. New York: Barnes and Noble, 1949.

Smith, L. M., and Geoffrey, W. Toward a model of teacher decision-making in an urban classroom. Cooperative Research Branch. Project 5-048, 1965.

Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarfeld, P. F., Starr, Shirley A., and Clausen, J. A. The American soldier: Measurement and prediction. Princeton, New Jersey: Princeton University Press, 1949.

Thibaut, J. W., and Riecken, H. W. Some determinants and consequences of the perception of social causality. Journal of Personality, 1955, 24, 113-133.

Young, P. T. Motivation and emotion: A survey of the determinants of human and animal activity. New York: Wiley, 1961.

PART II

Subject Matter, Content and Sequence

CURRICULUM improvement in the various disciplines has been hampered by lack of adequate research paradigms to determine the effectiveness of various instructional programs, by lack of relevant learning theory, and by inadequate conceptualizations for determining the content to be learned. The structure of the discipline itself, as formulated by one or more scholars, has provided much of the basis for recent curriculum development. E. G. Begle in "Curriculum Research in Mathematics" cites the extensive work of the School Mathematics Study Group in the development and evaluation of a mathematics curriculum. The structure of mathematics has proven to be a powerful organizer for the mathematics curriculum at the university, high school, junior high school, and now at the elementary school levels. The point of view adopted is that in learning mathematics each pupil is to build his own structure of mathematics. Such a structure starts with a small number of concepts and eventually incorporates strategies of problem solving. Begle argues that the necessary framework would be characterized by three aspects: "1) the mathematical content of the curriculum, 2) a theory of learning of the content, and 3) a theory of the application of the content to problem solving." Begle reports that some of the results obtained by the School Mathematics Study Group are being investigated in a National Longitudinal Study of Mathematical Abilities.

Similar problems of curriculum research and development are common to science. Arthur Livermore examines the structure of science (a science of science) as the basis for curriculum reform. Reports of the various curriculum groups (PSSC in physics, BSCS in biology, and CHEM and CBA in chemistry) indicate a union of scientists, educators and psychologists working together on the problem of curriculum design and evaluation. The evolution of the so-called new curriculum in science can be characterized as starting effectively first at the high school level, then junior high school, and now at the elementary school level. While these curriculum studies emphasize concepts of importance within the re-

spective disciplines, corresponding emphasis is also given to the empirical aspects of science as represented in the laboratory activities common to all of the curriculum studies. For example, the American Association for the Advancement of Science has produced a program for the elementary school which emphasizes the process approach to science. In this approach certain cognitive skills are used as the organizers of the curriculum. However, the lack of an encompassing conceptual framework in the science inhibits the development of a systematic curriculum K-12. Rather several well-prepared curricula are being developed at the various school levels.

Curriculum development in English is also dependent on the structure of that discipline. Robert Pooley discusses various efforts at developing curriculum rationales and materials in such areas as literature, composition, language, and in the emerging grammars of the English language. Development and research underway in the various areas are reported with particular emphasis given to the structural and transformational grammars.

Reading ability is of fundamental concern in the usual curriculum. The demands placed on the learner vary across age, and across subject matter, yet the student is not trained in techniques of reading science, mathematics or English. A research paradigm appropriate to investigating variation in the rate of reading is reported by Wayne Otto, Theodore Harris, and Thomas Barrett in "Research in Reading." The authors investigate the question of "whether variability in reading speed developed by training with carefully constructed, short model paragraphs will transfer to reading the less tightly structured passages typically found in classroom material." Extensive details on procedures and results are reported.

The contribution of psychological research to the development of curricula in mathematics is expanding, according to Myron Rosskopf. He cites the work of various psychologists in discussing the attainment of mathematics concepts. The effect of the learner's system of focusing on the problem at hand and the available strategy of problem solving are discussed.

Curriculum Research in Mathematics

E. G. Begle
Stanford University
Stanford, California

I SHALL interpret "curriculum" in the above title very broadly, since I believe that research in mathematics education must involve research on the learning of mathematics, the teaching of mathematics, problem solving, etc., as well as the syllabus, and that exclusive concentration on any one aspect of this complex is not fruitful.

I shall not attempt a historical survey of research in mathematics education. Weaver (1957-1966) has provided useful lists covering research on elementary-school mathematics for each year since 1951. Brown (1958, 1960, 1963, 1965) has provided biennial annoted reports of research on mathematics teaching carried out since 1956. A glance at these reports indicates that a considerable amount of research has been done in the past and also that the volume of research is increasing rapidly. A more careful look, however, provides the melancholy information that this large number of research efforts yields very little that can be used to improve mathematics education. This is not to say that all these efforts have been wasted. There are in fact many ingenious ideas and suggestions which ought to be followed up. In general, though, the results are too special or incomplete to be of any wide use.

It does not take much reflection to see why this should be the case. Research in education in general, and in mathematics education in particular, is much more difficult than research in say, psychology or chemistry. For one thing, the number of variables that should be taken into account is very large. Mathematics achievement can be affected, at least in theory, by the prior learnings of the subjects, by a wide variety of cognitive abilities and cognitive styles, by attitudes toward mathematics, and, as we well know, by socio-economic and demographic variables. Similarly the cognitive and affective state of the teacher may well affect the student's learning of mathematics.

In order to isolate and compare two different curriculum treatments, for example, it is necessary to take all these variables into account so that the effects due to these variables will not be confounded with those due to the curriculum treatments. To do this generally requires either a large experiment or a large number of carefully correlated small ones.

Until very recently, it has been almost impossible to meet this requirement. Experiments in mathematics education have been carried out by individuals with limited resources. The experiments of necessity were small in scale, involved relatively small numbers of subjects, and had a short time span. Only very recently has financial support become available which makes possible some of the studies and experiments necessary for significant improvement in mathematics education.

A large scale investigation now being carried out by the School Mathematics Study Group is the National Longitudinal Study of Mathematical Abilities (Cahen, 1965).

In this study three groups of students entering grades 4, 7, and 10 in September, 1962, totaling over a hundred thousand, will have been followed for five years. The mathematics achievement of each of these students has been measured twice each year.

The primary purpose of this study is to ascertain the effects of the new mathematics curricula in order to provide guidance for further curriculum development

It was necessary, of course, to measure not only the level of mathematics achievement of each student at the beginning of the study, but also as many as possible of the other variables which might affect the results. We have therefore obtained measures of a considerable number of cognitive variables, selected by our psychological consultants, for their probable relevance to mathematics achievement. An extensive attitude inventory was administered three times during the study, since we want to know how attitudes affect achievement and also how achievement affects attitudes. We have some information about the socio-economic status of the students as well as information about the schools. Most of the teachers of the students included in the study have provided us with extensive information about their training and experience, and more than half of them have also provided information about their attitudes toward mathematics, teaching, etc.

When we analyze this enormous mass of information, we will not only find out what we want to know about the differential effects on mathematics achievement of modern curricula as compared with conventional ones, but we will also obtain as a by-product, information on the effects on mathematics achievement of the other variables we have measured.

This kind of information should prove valuable both to the design of future experiments and in the interpretation of past results. Hopefully we will find that many of these variables have little or no effect on mathematics achievement. When this is the case, future experiments involving similar students need not take such a variable into account, and we will have more confidence in generalizing the results of past studies which did not control on such a variable.

However, while studies such as this one will prove very helpful, it is doubtful that they will prove sufficient. There is another fundamental problem which must be overcome before we can expect significant improvements in mathematics education. This problem is the lack of any theoretical framework for mathematics education. Without such a framework, one whose broad outline has reasonably widespread acceptance, there is no way to program the many individual small studies which will continue to be undertaken so that they can be mutually reinforcing and can contribute to overall forward progress. It is my belief that the time is ripe to begin the construction of such a theoretical framework.

THE MATHEMATICAL CONTENT OF THE CURRICULUM

One aspect of it of course will be concerned with the mathematical content of the curriculum. Before discussing this, I will turn to what may seem at first to be a digression. I am sure that many of our teachers look on our new curricula as a revolution in school mathematics. In a sense this is correct, but it is merely a small aspect of a revolution in mathematics itself which has been going on for a century and a half.

This revolution stems from the work of Abel and Galois. When Abel demonstrated the nonexistence of an algebraic formula for the solution of the quintic equation and when Galois applied his methods to settle the general problem, the direction of the development of mathematics was changed forever. Obviously they did not obtain their results by time-honored computational procedures. Instead they demonstrated that an examination of the over-all structure of a mathematical system, in contrast to computations with the individual elements of the system, was a very powerful mathematical tool and could lead to solutions of problems not otherwise solvable.

Although the revolution dates back to the previous century, it gathered momentum slowly, and it was only in the present century that its effects became overwhelming. The tremendous flowering of modern algebra and of functional analysis which started early in this century could not have taken place without the ideas about structure first pointed out by Abel and Galois.

During the 1930's these ideas began to be applied in the graduate curriculum, and it became clear that these ideas, already demonstrated to be powerful and indispensable in mathematical research, were also powerful and effective in mathematics instruction at this level. During the next two decades a start was made on a revision of the undergraduate curriculum in the universities along the same lines.

Consequently, when mathematicians joined with high-school teachers in an attempt to improve the secondary-school curriculum, their first inclination was to apply to the subject matter normally taught in these schools the same point of view toward mathematics, an emphasis on structure, which had proved successful not only in mathematical research but also in mathematical education at the university level. When it was discovered that this could be done successfully for the secondary-school curriculum, the reform movement continued and revised the elementary-school curriculum in the same spirit.

There are two important observations which now need to be made. The first is that this revolution has been successfully concluded. The second is that no new revolution is clearly in sight. Even if the first stirrings of a new revolution might be taking place in mathematical research, its effects could not appear in the pre-college program for generations. Consequently we can agree on the broad outline of the content of the mathematics curriculum for the schools.

This content is well enough known that we need not spell it out here in detail. Suffice it to mention that the mathematical topics included in the curriculum are, with few exceptions, those included in the pre-revolutionary curriculum. Only a few new topics, such as inequalities, have been added because of their intrinsic importance, and only a few other topics have been given lesser emphasis than before.

On the other hand, the new curriculum differs radically from the old in that it includes the structure of the common mathematical systems, the basic mathematical concepts and their interrelationships, as well as the basic mathematical facts and techniques.

Before discussing the implications of this new aspect of the curriculum for our theoretical framework for mathematics education, it is well to point out that while few new topics have been incorporated in the curriculum so far, there will undoubtedly be new ones to be added from time to time. The nature of our society at present is such that it would be wise to include as soon as possible, and for all students, some of the basic ideas of probability and statistics as well as some of the mathematical ideas relevant to high speed computers. Incorporation of any such new topic into the curriculum will require research to determine the prior understandings needed by the student, the most efficient location of the topic in the sequence, etc. We can certainly look forward to a continuing need for this kind of research.

Returning now to the curriculum, we see that our aim in mathematics education is to help each student to build in his mind a conceptual structure. This structure, for primary-school children, starts with a small number of concepts, abstracted directly from the real world, about number and geometric configurations. As the child progresses through school, these concepts are refined and sharpened, and the relationships between them became increasingly complex. At some stage, new concepts are met which come not from the real world but rather from previously learned concepts. The concept of irrational number is a case at point. Among these concepts there are, of course, those concerned with computation. For example, we do not ask our students to memorize an infinite multiplication table. Rather, we let them stop after 9×9 and teach them the algorithm for multiplication, and this is a concept.

Naturally, the order in which mathematical concepts are learned is subject to some variation, and much research time could be devoted to studying the effects of variations in the sequence. However, in my opinion, most of this time would be more usefully employed if devoted to some of the problems discussed below.

A THEORY OF THE LEARNING OF THE CONTENT

What this means for us is that our theoretical framework for mathematics education must have as one of its many aspects a theory of the learning of mathematical concepts and relations between these concepts.

No well-developed theory of this sort now exists. One's first thought, when the topic of learning of any sort is brought up, is to turn to our colleagues in psychology for advice and assistance. A very recent survey of concept growth in children (Wallace, 1965) shows that psychologists can provide very few useful answers to our questions. Much work has been done in this country on the development of concepts, but interest is usually focused on the strategies used by the subjects in trying to guess which combinations of color, sizes, shapes, etc., the experimenter has chosen. Neither the concepts nor the strategies have much relevance to classroom learning.

An interesting by-product, however, of these studies is that the availability of verbal labels for a concept plays an important role (Kuenne, 1946) and will have to be taken into account in any theory of the learning of mathematical concepts.

Some of the investigations of "discovery" teaching in this country have involved the learning of some non-trivial mathematical concepts. Unfortunately, the results of these studies present a confused picture. For a review of the present state of affairs, see Cronbach (1966).

Outside the U.S., investigations of concept learning have been dominated by the work of Piaget. He deserves great credit, of course, for investigating real mathematical concepts and how children come to acquire them. Unfortunately, he generally confines himself to the initial concepts of any branch of mathematics and has little to tell us about the learning of the later and more refined concepts or the learning of relationships between concepts.

A more fundamental difficulty in applying Piaget's findings in mathematics education lies in the fact that, when he questions a child to see how well he understands a concept, he is very careful never to provide the child with any subsidiary concepts which might help him. In fact, he seems to be investigating what children can learn without the benefit of education. There are, however, some indications (Kohnstamm, 1965) that the picture is quite different if children are provided with some assistance in the form of subsidiary concepts.

In short, the task of building a theory of the learning of mathematical concepts still lies before us. Some ideas about what will have to be included in such a theory can be obtained from the sources just mentioned. It should also be mentioned that some very interesting work, with which we are just beginning to become acquainted, is being done in Russia, and this work may provide us with further clues.

In order to construct a suitable theory of the learning of mathematical concepts, and such a theory should be quantitative rather than purely qualitative, it will be necessary to construct instruments which measure how well a student understands a particular concept.

This will not be an easy task. Concepts are not observable. It will be necessary to use indirect methods to assess a student's understanding of a particular concept. Perhaps we can obtain some guidance on this from our colleagues in the physical sciences, who also deal with unobservables such as neutrinos.

Perhaps the most useful contribution which Piaget has made to our present efforts is his demonstration that skillful questioning can provide precise and detailed information about an individual's understanding of a concept or problem. Similar individual interrogations will undoubtedly be an important tool

in the development of instruments for measuring understanding, although the final instruments will need to appear in a more standardized format. It is worth observing that Brownell (1963) also has found working with individual children a valuable procedure.

THE APPLICATION OF THE CONTENT TO PROBLEM SOLVING

A final aspect of our theoretical framework for mathematics education is that of problem solving. It would be futile to provide our students with a thorough understanding of mathematics if they were not able to use this understanding to solve problems. What we need therefore is a better understanding of how we can help our students to develop useful strategies for solving problems.

This seems to call for a considerable amount of exploratory work. We do have already, however, a variety of problem-solving strategies that can be investigated. Polya (1945) for example, has written extensively on this topic. In another direction, Newell, Shaw, and Simon (1959) have investigated, by means of computer simulation, the power of certain specific strategies. These strategies had been identified as ones actually used by at least some students, and a computer program then verified that the strategies were indeed effective. In still another direction is the very interesting work by Crutchfield and Covington (Crutchfield, 1966). Using comic-book format for their materials, they were able to train their subjects to examine the relationships between the various aspects of a problem situation, and the implications of these relationships, for a sufficiently long time and in sufficient depth to arrive at a solution. Their problems were not mathematical, but their procedures may well have implications for mathematics education.

The Russian literature, mentioned above, also has suggestions of problem-solving strategies which are worth investigating.

There are still two other sources which may provide us with raw material for investigation of problem solving. The strategies used by Newell, Shaw, and Simon were obtained by having a number of students solve problems aloud during individual interviews. This procedure may well prove very fruitful, especially at the elementary school level, where we seem to know less than at the higher levels. Similarly, classroom observations of teachers teaching problem solving and of their techniques for helping students to solve problems may pay good dividends.

In short, we already know a number of problem-solving strategies and we have suggestions for locating others. What we need to do now is to start a systematic study of techniques for teaching these strategies to students and of the ability of students to use these strategies. We need to study in particular the degree of understanding of the conceptual structure of mathematics a student needs before he can learn a particular strategy and use it effectively. This points up again the need for instruments which can be used to measure understanding of mathematical concepts.

To summarize briefly, I have been arguing that substantial improvements in mathematics education will have to be based on more fundamental research than has been carried out so far. I have further argued that the magnitude of the task is such that the efforts of a large number of research workers will be needed and that these efforts can be made mutually reinforcing and contributive to substantial improvement only if there is an overall theoretical framework which is generally accepted and which can be used not only to suggest individual research problems but also to relate individual findings in a meaningful pattern.

Three aspects of such a theoretical framework were mentioned: the mathematical content of the curriculum, a theory of the learning of the content, and a theory of the application of the content to problem solving. The first is, except for minor variations, dictated by the nature of mathematics itself. It differs from the curriculum content of a generation ago by emphasizing the conceptual structure of mathematics as well as routine mathematical skills. This part of our theoretical framework, then, is generally agreed to. The rest however, we will have to construct. We have a few bits and pieces to start with, but a great deal of empirical work will have to be done before even the outlines of these aspects of our general theory will emerge.

Finally, I have argued that an essential tool that we must have to carry out our task is a battery of instruments which measure understanding of mathematical concepts, and hence that our first step should be that of starting to construct such instruments.

CONCLUSION

In conclusion, I should like to note that any framework we may construct for mathematics education may be quite specific and not immediately generalizable to other subject matter areas. In mathematics we deal with only a very few aspects of the real world, namely the numerical and geometric aspects of physical situations. In science however, for example, many more different aspects of reality are considered, e.g., life, energy, mass, acceleration, color, density, galaxies, elasticity, genetics, the periodic table. However, as a student progresses through the curriculum, his mathematical concepts are refined, sharpened, and generalized to a much greater extent than his scientific ones, so the resulting mathematical conceptual structure differs markedly from his scientific one. Hence the process of concept formation in mathematics may not be the same as in science, and problem-solving strategies may have quite different values in the two areas. And when we move to more remote curriculum areas, such as social science or literature, for example, the situation is probably even more different.

REFERENCES

Brown, K. E. Analysis of research in the teaching of mathematics: 1955 and 1956. Bulletin No. 4. Washington, D. C.: U.S. Office of Education, 1958.

Brown, K. E., and Kinsella, J. J. Analysis of research in the teaching of mathematics: 1957 and 1958. Bulletin No. 8. Washington, D.C.: U.S. Office of Education, 1960.

Brown, K. E., Kinsella, J. J., Fehr, H., Myers, S., and Weaver, F. Analysis of research in the teaching of mathematics: 1959 and 1960. Bulletin No. 12. Washington, D. C.: U. S. Office of Education, 1963.

Brown, K. E., and Abell, T. L. Analysis of research in the teaching of mathematics. Bulletin No. 28. Washington, D. C.: U.S. Office of Education, 1965.

Brownell, W. A. Arithmetical abstractions – progress toward maturity of concepts under differing programs of instruction. The Arithmetic Teacher, 1963, 10, 322-329.

Cahen, L. S. An interim report on the national longitudinal study of mathematical abilities. The Mathematics Teacher, 1965, 58, 522-526.

Cronbach, L. J. The logic of experiments on discovery. In L. S. Shulman and E. R. Keislar (Editors), Learning by discovery: A critical appraisal. Chicago: Rand McNally, 1966. Pp. 76-92.

Crutchfield, R. S. Creative thinking in children: its teaching and testing. In O. G. Brim, R. S. Crutchfield, and W. H. Holtzman (Editors), Intelligence: Perspectives 1965. New York: Harcourt, Brace and World, 1966. Pp. 33-64.

Kohnstamm, G. A. Developmental psychology and the teaching of thought operations. Paedogogica Europaea, 1965, 1, 79-97.

Kuenne, Margaret R. Experimental investigation of the relation of language to transposition behavior in young children. Journal of Experimental Psychology, 1946, 36, 471-490.

Newell, A., Shaw, J. C., and Simon, H. A. Report on a general problem-solving program. Paris, France: Proceedings of the International Conference on Information Processing, UNESCO House, 1959.

Polya, G. How to solve it. Princeton, New Jersey: Princeton University Press, 1945.

Wallace, J. G. Concept growth and the education of the child: A survey of research on conceptualization. Occasional Publication Series No. 12. Slough, U. K.: National Foundation for Educational Research in England and Wales, 1965.

Weaver, J. F. Six years of research on arithmetic instruction: 1951-1956. The Arithmetic Teacher, 1957, 4, 89-99.

Weaver, J. F. Research on arithmetic instruction – 1957. The Arithmetic Teacher, 1958, 5, 109-118.

Weaver, J. F. Research on arithmetic instruction – 1958. The Arithmetic Teacher, 1959, 6, 121-132.

Weaver, J. F. Research on arithmetic instruction – 1959. The Arithmetic Teacher, 1960, 7, 253-265.

Weaver, J. F. Research on elementary-school mathematics – 1960. The Arithmetic Teacher, 1961, 8, 255-260. (a)

Weaver, J. F. Research on elementary-school mathematics – 1960 (concluded). The Arithmetic Teacher, 1961, 8, 301-306. (b)

Weaver, J. F. Research on elementary-school mathematics – 1961. The Arithmetic Teacher, 1962, 9, 287-290.

Weaver, J. F. Research on elementary-school mathematics – 1962. The Arithmetic Teacher, 1963, 10, 297-300.

Weaver, J. F. Research on elementary-school mathematics – 1963. The Arithmetic Teacher, 1964, 11, 273-275.

Weaver, J. F. Addenda to the May, 1965 report on research in elementary school mathematics. The Arithmetic Teacher, 1965, 12, 577-578. (a).

Weaver, J. F. Research on elementary-school mathematics – 1964. The Arithmetic Teacher, 1965, 12, 382-387. (b).

Weaver, J. F. Research on mathematics education, grades K-8, for 1965. The Arithmetic Teacher, 1966, 13, 414-427.

Curriculum Research in Science

Arthur H. Livermore
American Association for Advancement of Science
Commission on Science Education
Washington, D. C.

DEVELOPING and testing new science programs for grades K through 12 has been an activity of hundreds of university scientists and school teachers during the past ten years. The widespread and large-scale revolution in pre-college science teaching has been possible because large sums of money have been available from the National Science Foundation to finance the activities. The early curriculum reforms came at the high school level and these are now well established in school systems throughout the country. Because they are well known I will not elaborate on them, but it is important to review them briefly because they have, in large measure, set the pattern for science curriculum research in the pre-high school grades.

HIGH SCHOOL CURRICULUM RESEARCH

These high-school curriculum projects did several important things. In the first place they stripped away the authoritarian discussions of modern technological advances which had been tacked on to science courses that were twenty years or more out of date. They adopted the philosophy that teaching science was much more than teaching a catalog of facts. Rather, it should be teaching facts together with concepts and developing at the same time an understanding of and an ability to engage actively in the scientific enterprise.

To accomplish these things the approach to teaching was radically changed. No longer was the discipline presented in lecture and in text in didactic fashion. No longer did the laboratory serve as a place where the student demonstrated to himself facts already known. Instead, the laboratory has become the place where facts are learned through investigation and the classroom the place where the results of laboratory investigations are discussed and an understanding of fundamental scientific concepts are developed.

The Physical Science Study Committee (PSSC) led off the revolution in 1956. This group decided to deal with physics as a unified story in which the successive topics would lead toward an atomic picture of matter, and along with all this they hoped to develop the idea that physics is not a book completed or closed but is an unfinished and continuing activity. They designed new and simple laboratory equipment and devised new experiments so that the PSSC laboratory became the place where students carried on investigations, not simply demonstrations.

In a similar way, the other high-school curriculum groups pointed their courses in new directions.

The Chemical Bond Approach (CBA) project grew out of a conference of college scientists and high-school teachers held in the summer of 1957. The group noted that little change had taken place in high-school chemistry courses for at least twenty years and that modern chemical concepts were little mentioned and little used. The CBA group decided to make the most fundamental of all chemical concepts the central theme of the program. Hence, the idea of chemical bonds has become the unifying thread of the curriculum. Chemistry is presented as an interplay of experiment and ideas – of what is observed and what is thought. In the laboratory program the

students carry out small investigations, and in the classroom they learn to apply chemical concepts to explain their observations.

The Chemical Education Materials Study (CHEM Study) also stresses observation and concepts. This study grew out of the recommendations of an ad hoc committee set up by the American Chemical Society to consider needed changes in high-school chemistry. The philosophy of the CHEM Study group is that the important concepts and generalizations of chemistry should be developed inductively and that the evidence that the students use should be gathered by them in the laboratory wherever possible. Where this is not possible it is presented by teacher demonstrations or films. There is emphasis on the unifying concepts in chemistry rather than on great volumes and varieties of technical applications.

The Biological Sciences Curriculum Study (BSCS) of the American Institute of Biological Sciences faced some problems that were more touchy than those faced by the physicists and chemists. Traditionally, the topics of evolution and of human reproduction had been taboo in high-school biology courses. The BSCS group felt that modern biology could not be taught without introducing these topics, and they, therefore, included them in the three courses which they developed. Though there has been some resistance to these controversial topics in a few places, there is now general acceptance of the new courses by schools throughout the country. Indeed, it is estimated that this year 80 percent of all students taking biology in the tenth grade are in one of the BSCS courses. Though the three BSCS programs take different approaches to the teaching of biology, the goals are the same – to present biology as a dynamic and growing field of knowledge, not as an encyclopedic catalog of facts. A major accomplishment of the BSCS group has been to change the high-school biology laboratory from a place where dead, preserved specimens are dissected to a place where living organisms are examined and experimented with.

In every case the curriculum groups saw to it that the materials were thoroughly tested in classrooms before being released for general use. Each of the groups prepared batteries of tests to determine whether high-school students could, indeed, learn science in these new ways. The results have been gratifying.

One of the important accomplishments of all of the curriculum groups has been to bring together scientists and teachers to work out the new programs. These are not courses handed to the schools by college professors, but are truly programs that have been developed cooperatively.

Physics has never been a popular subject for high-school students, and the introduction of PSSC physics has not stimulated more students to study this discipline. There has been concern among scientists and educators that college bound students, in general, should have some experience in this field. One group having this concern is developing a new high-school program which, it is hoped, will encourage non-science majors to take physics. The program, called Harvard Project Physics, adopts a cultural approach. It develops important concepts such as force, energy, and motion, but draws heavily on the humanistic background of the science. The program is still in an experimental stage, and it remains to be seen whether it will achieve its goal of increasing physics enrollment in high schools.

The immediate short-range goals of the completed high-school curriculum projects seem to have been reached in a satisfactory manner. Through tryout and evaluation it is clear that high-school students can learn science through the medium of the new programs. There are some areas, however, where further research remains to be done. Each of the curriculum groups either explicitly or implicitly suggest that their approaches should develop in students an understanding of and an ability to engage in the scientific enterprise, and an interest in and enthusiasm for science which should carry on beyond the high-school years. Little has been done to determine whether the programs do indeed have this effect. A long-range study to evaluate this aspect of the new curriculum programs would be desirable.

Another area which should be investigated is the effect of one program on another in a school. For example, one might ask whether a student in PSSC physics in the twelfth grade does better if he has previously had BSCS biology and CBA or CHEM Study chemistry than if he has not. The AAAS Commission on Science Education has plans for carrying out such a study but the plans have not been implemented.

A third and important question is one of sequence of the courses in high school. The curriculum groups designed their programs to fit into existing slots: biology, tenth grade; chemistry, eleventh grade; and physics, twelfth grade. This sequence of courses has become established over several decades but there is no logical reason for teaching the courses in this order. Indeed, when one looks at the interrelationships of the disciplines, logic suggests that the order should be inverted. Modern chemistry is deeply rooted in physics, and modern biology requires of the student more than a speaking acquaintance with chemistry. The groups who prepared the chemistry and biology curricula recognized the problem. The chemists were forced to include some of the ideas of physics - for example, the concept of energy -- in their programs. The biologists in turn, had to include some simple chemistry in the hope that students would better understand some of the molecular biology which is included in the courses. The result in each case is not completely satisfactory, and it seems clear that some research needs to be done on this problem. I shall deal with this later. But now I want to turn to the pre-high school years and discuss the development of science curricula there.

PRE-HIGH SCHOOL

I said earlier that the new high-school science programs have set the pattern for science curriculum research in the pre-high school years. The pattern was set in several ways, including cooperation of scientists and teachers in producing the new programs, and emphasis on student activity and

and investigation rather than on didactic presentations. As the new high-school curricula came into being, the writers and the teachers who used them soon realized that many of the skills and attitudes that students need for effective participation in the new curricula should be learned in earlier grades. Let me take just one example. The scientists working with high-school teachers developing CBA chemistry soon learned from the teachers that they could not assume that students entering high-school chemistry had developed skills in constructing and interpreting graphs and so a section of the CBA program is devoted to developing these abilities. There seems to be little doubt that these are skills that could be developed in earlier grades.

Many of the scientists who worked on the high-school curricula are now devoting their energies, in collaboration with school teachers, to designing and testing science curricula for the early grades. There are so many of these programs for elementary and junior high school that we cannot examine them all in detail. But we will review most of the major ones starting with the programs for the early grades and telling you in some detail about the elementary science program on which I am working.

Each of the groups working at the elementary school level have the same general philosophy and plan of attack. The purpose of introducing new science programs into the early grades is not so much to recruit more scientists as to prepare individuals to be scientifically literate citizens. Literate not in the sense of knowing many scientific facts, but of understanding both the power and the limitations of science and its fundamental concepts.

The approaches that the various curriculum development groups are taking to achieve this goal vary widely. The Elementary Science Study of Educational Services, Incorporated takes the position that the child should be introduced to science by presenting him with a variety of materials to investigate. They have developed and tested in classrooms, units on butterflies, gases and airs, kitchen physics, microgardening, meal worms, and many other topics. Let us look at the meal worm unit as an example.

The children are given shoe boxes containing a few meal worms and are left to investigate them on their own. Gradually, as they observe them, they begin to ask questions. How do the meal worms find the edges of the box? How do they find their food? How do they back up?... and so on. As they design experiments to answer their questions, it is hoped that they will learn to evaluate evidence based on observation, classification, measurement, and control of variables; that they will learn to ask questions that they can answer by simple tests; and that, at the same time, they will develop an enthusiasm for scientific investigation.

The School Science Curriculum Project of the University of Illinois has a philosophy similar to that of ESS.

The Science Curriculum Improvement Study at the University of California takes a different approach. Its main thrust is to develop an understanding of science concepts starting in kindergarten with the simple concept of "What is an object?" and moving from there to "What is a system?", "What is meant by interaction?", and so on. Their program is more structured than ESS but still leaves room for investigation by children.

A science program for the early grades is the Minnesota School Mathematics and Science group, abbreviated "Minnemast." This group began by developing a mathematics program and then devising science units to go along with it. Their science program aims at developing both skills and concepts. So far they have seven units for kindergarten and first grade entitled "Watching and Wondering," "Describing and Classifying," "Our Senses," "Shape and Symmetry," "Objects and Their Properties," "Changing and Unchanging Properties," and "Introduction of Measurement."

The project with which I am working is developing a program, called Science – A Process Approach, for K through grade 6. This curriculum development is one of the activities of the AAAS Commission on Science Education. Shortly after it was established in the spring of 1962, the AAAS Commission on Science Education held two conferences of scientists, science educators, and teachers. From these conferences came the suggestion that a fruitful approach to teaching science in the early grades would be to devise exercises which would develop children's skills in using the processes of science. The Commission elected to carry out this task. This program, which is now being tried out in schools for the third year, is, like the other projects, a cooperative effort of scientists and school teachers.

A basic premise of this project is that a scientist's behaviors as he engages in his pursuit constitute a complex set of skills and intellectual activities but that these can be analyzed into simpler activities, and it is these simpler activities which we are calling the processes of science. Another premise is that the individual's ability to use these various processes can be developed in a step-wise fashion beginning in the earliest years of school. The processes for the early grades have been identified as observing, classifying, measuring, communicating, inferring, predicting, recognizing and using space/time relationships, and recognizing and using number relationships. For each of the processes there is a series of exercises of increasing complexity and difficulty which require the child to engage in activities in the classroom rather than to simply read or be told about science.

Let me give you a better idea of what the program is all about by describing in some detail one of the exercises written for the first grade. The exercise is the fourth one in the Communicating process and is entitled, "Introduction to Graphing." It starts with a statement of objectives – "Following this exercise the child should be able to":

1. construct a bar graph.
2. identify and name the number of items represented by the bars of such a graph.
3. state the number of items represented by the bars and distinguish between such expressions

as more than, fewer than, the same number as, most, and fewest.

There is a section called Rationale which tells the teacher why graphing is introduced at this level. It gives her some suggestions about contructing bar graphs and also includes suggestions of how to handle children's questions about why a graph is useful in communicating information.

The activities are introduced by putting into a box a number of colored blocks so that there will be one for each child and so that there will be at least three different colors. The teacher then asks each child to come and pick a block out of the box and then put it into a rough pile on the table so that the blocks of each color will be piled together. The teacher then asks the children if it is easy to find how many blocks there are of each color. The children, of course, find it necessary to come up to the table and count the blocks in the piles. The teacher then asks if they can think of a way of arranging the blocks so that the number can be found more easily and suggests, if the children do not think of it, that if the blocks are piled in vertical columns, they can be counted easily even from a distance. This leads ultimately to the question of ways of remembering how many blocks there are of each color and into the next activity in which bar graphs are first introduced. In that activity each child again selects a block and also a felt square and as he puts the block in one of the piles on the table he puts the felt square in one of the columns on a felt board. At the end of this activity there is a bar graph consisting of felt squares on the felt board, one column for each color of block. In the activities that follow the children make bar graphs by coloring squares on large sheets of paper and finally are introduced to the idea of a grid and of labeling the vertical axis of the grid with numerals and the horizontal axis with a key to identify the objects.

At the end of this exercise is an Appraisal which the teacher uses to determine whether the children have developed the skills set forth in the objectives. For the Appraisal the children are given objects such as toy vehicles of various sorts and asked to make graphs to indicate how many of each type there are and then to answer questions the teacher poses about the graph – questions such as: Are there fewer cars than motorcycles? Are there more trucks than motorcycles?

You will have noted that one of the processes that I mentioned is recognizing and using number relations. While there were some of the writers who felt that it was not necessary to include number exercises in a science program, the majority agreed that since number skills are required in many of the other process exercises, we should not simply assume that the child learned them in his regular mathematics program. The exercises in the numbers process are not intended to be a complete mathematics program but are intended to supplement the program already in use in the classroom.

At the fourth grade we begin to shift gears and introduce a new set of processes which are more clearly recognizable as activities that a scientist engages in. These processes are formulating hypotheses, controlling variables, experimenting, interpreting data, devising models, and making operational definitions. We sometimes speak of these as the integrated processes because they bring together the skills from the earlier part of the program.

One of the advantages of a program in which the objectives are stated in behavioral terms is that evaluation can be done by assigning the children tasks which will indicate whether or not they have attained the expected skills. For example, following the kindergarten exercise in which the children learn to identify and to name various plane shapes, such as circles, squares, triangles, and rectangles, it is a simple matter for the teacher to present a child with an array of plane shapes and ask him to point to a triangle, or the teacher might point to a circle and ask the child to name it. The results of each task can be recorded with a binary scale. Yes, the child could perform the task, or no, he could not. For each exercise we have prepared a series of tasks which we call the competency measure. Each teacher in our tryout program uses the competency measures on a sample of children from her class after she has taught each of the exercises. The results of this testing in the various tryout centers are collected and tabulated and used by the writers as they revise the exercises during each summer writing session. The results give the writers a good idea about the success or failure of any given exercise.

There are many aspects of the teaching in the classroom that we cannot learn about through the competency measure and so we ask each teacher to give her subjective impression of the exercise on a feedback form. This includes information about the time required to carry out the instruction, suggestions of activities that might be omitted or added, the reaction of the class, including interest and enthusiasm, carry over of learning to other activities or areas, and attention span. This feedback is also helpful to the writing group as the program is revised.

Although the emphasis in Science – A Process Approach is on developing skills in using science processes it is important to ask the question, "What about science content?" It is obviously impossible to teach process skills without using content of some sort, and the content that we use is drawn from various fields of science. To insure this we include in our writing team scientists from all major areas. In the early grades, of course, there are many exercises in which the relationship of the content to science appears to be somewhat distant. Learning to recognize and name two-dimensional and three-dimensional shapes is an example. Learning to make graphs using colored blocks is another. However, as the program progresses into higher grades the content is more and more clearly science. For example, among the twenty-five exercises in Part Five, which is used in fourth grade, are the following:

Inferring 8 – Inferences of Patterns in Electric Circuits
Defining Operationally 1 – Circuits, Conductors, and Nonconductors
Controlling Variables 1 – Growth of Mold on Bread

Interpreting Data 1 - Guinea Pig Learning in a Maze
Controlling Variables 7 - Orientation of Plants
Experimenting 1 - Separating Mixtures

Of the eighty-one exercises in Parts Five, Six, and Seven, the subject matter is approximately as follows: physics 33 percent; biology 21 percent; mathematics 17 percent; chemistry 9 percent; psychology and social science 9 percent; earth science 7 percent; astronomy 4 percent.

JUNIOR HIGH SCHOOL

We believe, then, that by the end of the sixth grade, children going through this program will not only have developed sufficient skills to carry out simple investigations but also will have a store of scientific knowledge which will make it possible for them to embark on a fairly sophisticated science program in the junior-high-school years. What the pattern for junior-high-school science will turn out to be is not yet clear. At the moment there are four groups who have developed curriculum materials for junior high school and tested them on a fairly wide scale. Then there are several other groups who are at earlier stages of curriculum development. The four programs that are farthest along are centered mainly in the physical and earth sciences and astronomy.

The astronomy program, known as the University of Illinois Elementary School Science Project, actually is designed for the later elementary grades and the early junior-high-school grades. This program is frankly subject matter oriented and consists of a series of childrens' books and teachers' guides. The program starts with the topic "Charting the Universe," which includes measuring distances in the solar system, measuring the size and shape of the earth, and using light as a tool for measuring great distances. The topics which follow it are "The Universe in Motion," "Gravitation," and "The Message of Starlight." Two additional topics which are still in preparation are "The Life Story of a Star" and "Galaxies and the Universe." These materials are not intended to serve as a complete science program for grades five through eight, but rather as topics to be included in a broader science program. The materials are being tested by several hundred teachers in various parts of the country but the results of the evaluation have not been reported

The second program which includes some astronomy but many other topics as well, is the Earth Science Curriculum Project which is designed for ninth grade. This course was first written by a team of scientists and teachers in the summer of 1964, revised in the summer of 1965, and is already being prepared for commercial publication. The course, called Investigating the Earth, is an experience-centered course in which inquiry in the study of natural phenomena is stressed. It consists of four units - the first one, "The Dynamic Earth," considers the materials of the earth, earth changes, earth motions, force fields, and energy flow in earth processes. The second unit, "Earth Cycles," considers the hydrologic cycle, climate, and the earth's crust. The third unit, "The Earth's Past," discusses the measurement of geologic time and paleobiology. The course ends with a fourth unit, "Earth and the Universe," which begins with a discussion of the moon, moves on to the solar system and its origin, and finally to stars, galaxies, and cosmology. The program has been evaluated in eighty classrooms, and student progress has been determined with a test on science knowledge which is used both as a pre- and posttest.

A third junior-high-school program - Time, Space, and Matter - is being developed by the Secondary School Science Project at Princeton University. This program begins and ends in geology but includes a good deal of fundamental physical science as well. The students investigate topics such as solution, crystallization, and water erosion using well-designed equipment which is furnished to them in the form of individual kits.

A good deal of attention is being given to evaluating this program to provide feedback for those who are revising the course. Each teacher keeps a daily record of course progress and reports at regular intervals to the project office. The teachers are particularly requested to indicate where students show an extension of knowledge or interest to out-of-class situations and where they show signs of understanding or of confusion as they use the course materials. In addition, the group is developing a test to demonstrate changes in the behavior of children after they have had experience with the new curriculum. Other methods of evaluation, including the instructional interview technique, are being considered.

A fourth program, called Introductory Physical Science, is being developed by the same group that produced the high-school PSSC physics course. This course also is intended for ninth grade and is designed to develop in the students basic attitudes and skills that they will need when they later encounter the new high-school programs in physics, chemistry, and biology. The first two-thirds of the course considers matter and its properties and leads to the need to assume that matter is particulate in nature. The last third of the course considers the size of molecules, molecular motion, and energy. This course is laboratory centered, and the group has devised simple equipment which permits students to carry on significant experiments in ordinary classrooms.

It is clear as one looks at the new developments in science curricula for the junior high school that we are a long way from a smooth flowing science curriculum for these grades. There has been little or no coordination among the present curriculum development projects. As a first step toward coordination the AAAS Commission on Science called a conference of representatives of the junior-high-school curriculum projects in science and mathematics. The conference was held in July 1965 at Michigan State University. As the representatives of the projects reported on their work it became clear that there was a major gap in the junior-high-school curriculum developments. No group was working in the field of biology. The Biological Sciences Curriculum Study has, for some time, been planning to develop a biology program for junior high school, but they have made no progress since funds have not been made available to them.

At the conference, Bentley Glass and Arnold Grobman of BSCS suggested two possible types of junior-high-school science programs. One is a topical plan and the other a stratified plan. In the topical plan they suggested that the science program for grade 7 include about half a dozen topics, such as conservation of matter and energy, natural cycles, biological regulation, and homeostasis. Another half dozen topics would be developed for eighth grade. These would include change through time – cosmological, geological, and biological. Another group of topics would be developed for ninth grade and would include the broad aspects of meteorology, environment, and habitat.

The stratified plan that they suggested was as follows:

Grade 7 – physical science – a year course in descriptive aspects, largely earth sciences.
Grade 8 – life science – a year course in diversity and unity in the organic world, fundamental structure and function.
Grade 9 – physical science – a year course in basic concepts of matter and energy.

Another possible approach to an integrated junior-high-school science program was described at the conference by Dr. Robert Jastrow, chairman of the Junior High School Science Subcommittee of the AAAS Commission on Science Education. This program would follow naturally from the elementary program, Science – A Process Approach, but could also build on the other elementary science programs that I have already described. The course for seventh grade would serve as a transition between an elementary science program where the student has developed skills in using science processes and where he has learned some basic science content; and the eighth and ninth grades in which the course would develop a coherent picture of man and his physical and biological environment. The seventh-grade course, which would be rooted in laboratory experiences and field trips, would consist of selected topics from the physical and biological sciences and from astronomy and the earth sciences.

The eighth grade would begin with consideration of the basic materials and forces of the universe. The neutron, proton, and electron as fundamental units out of which all matter is constructed, and the basic forces of nature - gravity, electromagnetism, and the nuclear forces. The interplay of particles and forces would then be used in the discussion of the birth of stars and galaxies and the synthesis of chemical elements. The next step would be to consider current ideas about the formation of the solar system, nature of planets, and the probability that planets exist around other stars. Then geophysics and geology would be introduced in a discussion of the history and present structure of the earth and its atmosphere, and finally the course would take up the elements of molecular biology and combine this recently acquired biological knowledge with the geological discussion of the earth's history in broad attack on the problems of the origin of life and the biochemical foundations of evolution.

The ninth-grade course would treat biology from the point of view of evolution. The first topic to be considered would be the development of self replicating cells from primitive catalysts and complex organic chemical systems. Then multicellular organisms would be discussed and the similarities between plant and animal forms as far as capture, storage, and step-wise release of energy is concerned. Following this, differentiation and embryogenesis would be considered and then the molecular, organismic, and ecological factors involved in natural selection would be treated. The ninth-grade program would end with a discussion of the development of the nervous system and with how the nervous system accounts for consciousness, perception, learning, and memory.

A course of this sort, in which central problems of modern science serve as the core, should serve as a general education course for students who do not elect to take much science in high school. At the same time, it should provide a solid background of the skills and understanding which would serve as an effective base for the high school biology, chemistry, and physics courses. No work beyond planning has been done on this program yet. It is hoped that work can begin within the next year.

SCIENCE COURSE SEQUENCE

Now I want to come back to the question about the sequence of science courses in high school that I have already mentioned briefly. One of the arguments for putting physics off until the twelfth grade is that the students do not develop the necessary mathematical skills until they have reached that level. I have not said anything about the new mathematics curricula except the relationship of mathematics to the Science – A Process Approach elementary curriculum. There are, of course, several new approaches to teaching mathematics in the schools. In these new programs the students develop mathematical skills at much earlier grade levels than they have done in the older traditional programs. It is possible, I think, that by the end of the ninth grade, students will have sufficient skill in algebra to do the mathematics in a modern physics course such as PSSC. Mathematics skill would, therefore, no longer be a block to moving physics to an earlier grade.

If a junior-high-school science curriculum gives the student some fundamental knowledge of physics, chemistry, and biology and related sciences, I see no reason why physics cannot be made a tenth-grade course, followed by chemistry in grade 11, and a molecular biology course in grade 12. Already, in a number of school systems around the country this sequence is being tried.

There is another possible pattern for grades 10 and 11. Since there are a number of topics that are common to physics and chemistry, it is possible to interweave these two disciplines into a two-year sequence of physical science. An experiment is now under way in Portland, Oregon, to do just this. A group there set out three years ago to devise a "road map" through PSSC physics and CBA chemistry and a similar one through PSSC physics and CHEM Study chemistry so that teachers could move easily from one book to the other. The program starts with the topic "Interpreting the Universe," and considers fundamental concepts and measurements and

the nature of chemical change. The next section is mechanics which is all drawn from the PSSC course. The rest of the year is devoted to the electrical nature of matter, atomic models, and molecules and energy. In all of these chemistry and physics are interwoven. In the second year of the program there is less blending of the two subjects. The course begins with optics and waves, followed by electromagnetics and a discussion of quantum systems and then moves into consideration of bonds and chemical systems; order, disorder, and change; and molecules and energy.

This program has been evaluated using selected test items from CBA, CHEM Study, and PSSC tests. The Portland group reports that the combined program is successful and that even though students are permitted to take only the first year if they like, a large majority of them stay with it for the full two years. At present, this program is designed for grades 11 and 12, but it could be used at grades 10 and 11.

The "road map" technique is not completely satisfactory, and there is need for a fresh start to develop an integrated modern physical science curriculum.

REVIEW OF PROBLEMS

We have examined some of the recent and current science curriculum research and have suggested directions in which curriculum research might go in the future. There are some corollary problems that I have omitted. We will review these briefly.

As elementary science education moves from fairly passive learning through books and exposition to student manipulation and investigation, there will be an increasing need to provide supplies and equipment for classrooms even as early as kindergarten. The amount of money provided for science supplies in most school systems is small. It will have to be increased manyfold to accommodate the new programs. It is impossible to do more than estimate roughly what the cost of introducing the new elementary programs would be. It appears that on the order of $200 per classroom will be needed to introduce the new programs. To assist school systems in planning for the introduction of the new elementary science programs, the AAAS Commission on Science Education is just embarking on a study of the costs of equipment and supplies. It is our expectation that this study will be completed shortly.

Classroom design will also be influenced by the new science curricula. As a minimum for the early grades, classrooms should have multiple electrical outlets, running water, and level working areas. I have been surprised at the many classrooms I have visited that have only a single electrical outlet.

Although the new junior-high-school science curricula that are now under development are designed to be used in classrooms without special laboratory facilities, I think that in the future junior-high-school classrooms will need to be provided with laboratory space similar to that found in senior high schools.

A crucial problem raised by the new curricula is teacher training. As you know, the senior-high-school science teachers have been able to attend summer institutes to learn to use the new biology, chemistry, and physics courses. A few institutes are now being held for junior-high-school science teachers. For the elementary schools the problem is different both quantitatively and qualitatively. The science programs for the earliest grades do not require that the teacher know a great deal of science content, but they do require some teacher orientation. Up to now during the experimental stage the various curriculum groups have carried on this orientation and training on an ad hoc basis. In the future, as the programs are more widely used, school systems must establish inservice training programs. It would seem that high-school science teachers might be used to staff such programs.

And then there is the problem of preservice training of teachers. Only one of the curriculum groups, "Minnemast," is preparing college material to train prospective teachers to use the "Minnemast" materials. There is a group at Rensselaer Polytechnic Institute that is designing a new course in physical science for non-science majors with the thought that this course will better prepare new teachers to use the new school science curricula.

I said earlier that one of the important things that happened as the curriculum groups organized and developed their programs was that scientists and teachers collaborated, each learning from and educating the other. This cooperation has produced needed and significant changes in science curricula. Changes which could not have been effected by either scientists alone or teachers alone. This cooperation must, and I am sure, will, continue. Science curriculum research is a continuing activity, and the best minds in science and the best minds in education must be devoted to it.

Curriculum Research and Development in English

Robert C. Pooley
The University of Wisconsin
Madison, Wisconsin

THIS PAPER is planned to present selected items of recent research in the teaching of English from a number of aspects. Together they make up the body of information and expert opinion which will largely direct curriculum work in the decades ahead. Although the goal of most English teaching is to accomplish an integration of the component parts of the subject, they are here presented in divisions to avoid confusion and overlap. One area of research and study often associated with the English language arts is reading as a skill. It is omitted from the survey because in the field of education it is considered separate from the language arts. It has its own courses, its own facilities, and a very extensive body of research, only a small portion of which is related to the language arts. The word "research" has been interpreted broadly enough to include mention of seminal books which, if not in themselves research, have induced and will induce research in the areas of their treatments.

THE PROFESSION OF ENGLISH TEACHING

In 1961 a committee of the National Council of Teachers of English published a document entitled The National Interest and the Teaching of English with the sub-title A Report on the Status of the Profession. In three major parts this document studied the present situation in the teaching of English, related the situation to the national scene, and made analyses and studies leading to specific recommendations for the improvement of the teaching of English by meeting recognized needs.

The scope of the responsibility of English is dramatized by a bar chart revealing that of all high school students in school in 1960, 92.9 percent were enrolled in English courses, 68 percent in social sciences, 55 percent in mathematics, and 25 percent in foreign languages. A survey revealed that in the years 1958-1959 the demand for trained high-school teachers outran the supply by 27 percent. Expanding school populations will increase this ratio. Moreover, the expansion of knowledge and the requirements of a complex society will increasingly demand more technical training and knowledge on the part of English teachers than was true in the past. The single greatest weakness in the current situation is the lack of articulation in the teaching of English from the elementary school through college.

A summary of the highlights of the status report follows:

The Need for Teachers of English:

a. The shortage of teachers shows no sign of decreasing.
b. One-fourth of all elementary teachers are not college graduates.
c. Only 40 to 60 percent of high-school English teachers have completed a college major in English.
d. There is an acute shortage of college instructors in linguistics and composition.

State Certification Laws:

a. States differ widely in requirements.
b. Ten states certify elementary teachers without a degree.
c. Nineteen states have no stated requirement in English for elementary teachers; twenty-

one lack definite requirements in reading, children's literature, and English methods.
d. Minimum preparation in English for English teachers is sixteen to eighteen credit hours.

Preparation of elementary teachers:
a. More than 94 percent of the colleges fail to require specific study of the English language.
b. Less than half require a course in the teaching of the language arts.

Preparation of secondary-school teachers of English:
a. Only one-fourth of the colleges require a course in the history of the English language.
b. Only 17 percent require a course in Modern English Grammar.
c. Only 41 percent require a course in advanced composition.
d. Only slightly more than half the colleges require a course in the methods of teaching English.

Better Teaching Conditions:
Teachers with a student load of 150 (five sections of thirty students) must work more than 50 hours a week just meeting classes and marking papers. (In 529 out of 682 schools, English teachers had five classes.)

Libraries:
a. 10,000,000 elementary, and about 150,000 high-school students attend schools lacking a central library.
b. The average expenditure for books per student is about half the price of one average book.

Remedial English in Colleges:
a. An estimated 150,000 students failed college English tests in 1960.
b. Two-thirds of American colleges have to offer remedial work in English.
c. The cost of remedial instruction is estimated to be over ten million dollars.

A continuation of the research into the status of the teaching of English was carried on by a committee of the National Council of Teachers of English and published in 1964 under the title The National Interest and the Continuing Education of Teachers of English. The principal purpose of this investigation was to discover and reveal the need for inservice education of teachers concerned with English in elementary and secondary schools. Questionnaires were answered by 7,417 secondary-school English teachers and by 3,030 elementary-school teachers.

Some highlights of the analysis are:
1. Over 40 percent of the nation's elementary teachers began full-time teaching without a degree. Only 31 percent of these achieved the degree after beginning teaching.
2. More than one in ten high-school teachers (11.8 percent) began full-time teaching without a degree. Practically all of these earned a degree later.
3. Although 24 percent of instructional time in kindergarten through grade 12 is given to English and language arts, only 8 percent of the elementary teachers' college preparation is given to this subject.
4. About half (49.5 percent) of secondary teachers teaching English lack a major in this subject.
5. By their own admission only 51.9 percent of secondary-school English teachers feel competent in literature; 36.6 percent feel competent in composition, and 53.5 percent feel competent to teach the English language.
6. In voluntary continuing education, elementary teachers take one summer session for each eight years of experience. But the courses taken are largely in education, with little or no further education in the language arts.
7. Thirty percent of secondary teachers of English have not taken a course in English in more than ten years.
8. Interest in further education for elementary teachers was expressed for six areas: children's literature, language, speech, writing, general literature, and the largest number, 79.1 percent in reading.
9. Secondary teachers expressed interest in further experience in practical methods of teaching English: intermediate or advanced composition, literature for adolescents, teaching of reading, literary criticism, and structural and/or generative grammar.

The broad conclusion is overwhelming evidence of inadequate training with a clear need for means to make inservice further education available to nearly all teachers of the English language arts.

Two books of importance to the whole profession of the teaching on English appeared in 1963 and 1965 respectively. The Education of Teachers of English (Grommon, 1963) is Volume V of a curriculum series of the National Council of Teachers of English. Written by a committee representing elementary-school, high-school, and college English teaching, it analyzes facts, implications, and inferences in such matters as the liberal education of teachers, the professional education of English teachers, and the specific preparation in English itself. It confirms the conclusions reached in other studies that elementary-school teachers are very poorly prepared to teach English; that the preparation of high-school teachers is subject to wide variations and some serious omissions; and that college teachers, though perhaps well grounded in content, are often innocent of any preparation to teach. Much space is devoted to recommendations and procedures for the continuing education of teachers by means of summer sessions, institutes, evening classes, and uses of mass media.

The second broad area book is Freedom and Discipline in English, a report of the Commission on English of the College Entrance Examination Board (1965). The Commission on English was appointed in September 1959. Consisting of sixteen members from universities, colleges, and private schools, it undertook to define the field of English, and to set standards for instruction in English and in the preparation of teachers for English. In 1962 it sponsored and supported twenty institutes attended by 868 teachers in all parts of the United States. It commissioned and issued a series of kinescopes featuring experienced college teachers of English expounding their specialties. These kinescopes are lent free of

charge except for mailing costs. The volume Freedom and Discipline in English is the final report of the Commission.

Chief contributions to the curriculum are:
1. Recommendations to improve the quality of instruction.
2. An analysis of language teaching, including grammar and usage.
3. An analysis of literature teaching, emphasizing the critical process in teaching and studying literature.
4. An analysis of composition, including valuable suggestions for motivating writing, and for the correction, criticism, and measurement of writing skill.

Without essential conflict in fundamental matters, it may be stated that, in general, the publications of the National Council of Teachers of English tend to be experimental and progressive in outlook, and that the point of view of the Commission on English is conservative. To the curriculum director there is great value in the balance to be derived from these differing viewpoints.

RESEARCH IN THE TEACHING OF LITERATURE

Although most of the studies in literature deal with ideas and values rather than with controlled evaluation, some recent studies reveal effective techniques for measuring the success of students in the learning of kinds of literary materials. Squire (1963) studied the responses of adolescent readers to four short stories, establishing their responses as forming seven categories: I. Literary Judgments: Direct or implied judgments on the story as an artistic work; II. Interpretational Responses: Reactions in which the reader generalizes and attempts to discover meaning, motivation, and the nature of the characters; III. Narrational Responses: Reporting facts or details from the story without interpretation; IV. Associational Responses: Responses in which the reader associates ideas, events, places, or people with his own experience; V. Self-involvement: Responses in which the reader associates himself with the behavior and/or emotions of the characters; VI. Prescriptive Judgments: Responses in which the reader prescribes a course of action based on some absolute standard; VII. Miscellaneous Responses: all others.

Wilson (1966) reports the responses of college freshmen to the reading of three novels: The Catcher in the Rye (Salinger), The Grapes of Wrath (Steinbeck), and A Farewell to Arms (Hemingway). Fifty-four freshman English students read a novel, took notes while reading, and wrote immediate reactions in class. After three discussion periods they wrote follow-up reactions. This procedure was followed for each of the three novels.

The written responses were coded by means of Squire's (1963) categories. Statistical analysis showed that 65.6 percent of responses to the three novels were interpretational; 12.2 percent literary judgment; self involvement, 8.9 percent; narrational, 8.7 percent; associational 2.1 percent; prescriptive 1.3 percent; miscellaneous 1.3 percent. The lead of interpretation over literary judgment is a sign of maturity of perception; it suggests that teachers should avoid supplying "correct" interpretations before the book is read by the student. The study reveals that although students may begin their involvement with literature in a groping and emotional fashion, their later attempts at interpretation become sharply formulated and logical. Teachers should therefore refrain from demanding that first reactions be fully described or defended. But as the study of the novel moves into the grappling with difficulties, the instructor should demand expert analysis. Most students are capable of growth of interpretation by these means.

The Research Foundation of the National Council of Teachers of English is sponsoring, directing, and supporting research in children's growth in sensitivity to, and appreciation of literature. Starting with a national conference of elementary-school, high-school, and college specialists in literature assembled in Chicago in May 1963, the Foundation applied itself to research in the teaching of literature because so little is being done. It was soon discovered that there were no instruments available to measure any aspect of success in children's growth in literature. The Foundation then determined to try to create such tests, both as research in itself, and as the means to further studies. Following consultations with various authorities, the Foundation appointed Margaret Early of Syracuse University the Director of the test activity and appropriated funds in its support. In the summer of 1965, Early assembled at Champaign, Illinois, some twenty-five selected teachers and librarians to develop test items. From the materials produced by this group, and augmented by Early's advisors, four forms of a test have been developed by cooperative work with the Educational Testing Service of Princeton, New Jersey. A trial run was made of these forms in the spring of 1967 at elementary schools in grade 5 in locations covering the United States. From the statistical study of the returns it is expected that two final forms of the test will be developed.

The goal is to attempt to discover growth in the responses of children to literary experiences. The range is from grade 4 to grade 6, inclusive. Although it will be a pen and paper test, many of the items will be presented orally on tapes. If the test reveals measurable growth in responses to literature, the research will continue in two directions: (1) to measure the effectiveness of kinds of instruction on the literary growth of children; and (2) to create similar tests for junior-high schools and possibly senior-high schools.

RESEARCH IN THE TEACHING OF COMPOSITION

One of the most exhaustive researches of recent times was conducted by Richard Braddock, Richard Lloyd Jones, and Lowell Schoer of the University of Iowa and published in 1963 by the NCTE under the title Research in Written Composition. It undertook to assess all known research in composition that could be described as "scientific." From twenty bibliographies of research came over a thousand citations; committee analysis reduced these to 485 items. From consultation with seventeen authorities

this number was reduced to about one hundred titles; later some fifty other titles were added. After vigorous analysis, only five studies survived the criteria set up by the investigators. These five only, in the eyes of the investigators, can be termed valid in the teaching of composition.

1. The Baxton study revealed that thorough criticism of composition by a competent reader, followed by revision by the student brought better gains than writing not criticized.

2. The Harris study showed in a 2-year space that the direct (or functional) teaching of grammar brought better results than the formal, analytical method.

3. The Kincaid study undertook to reveal whether a single paper written by a student on a given topic at a particular time provided a valid means to judge a student's writing ability. Joined to this question were several others, so that clear answers were almost impossible to secure because of weakness in the structure of the research.

4. The Smith study dealt with class size as a factor in the teaching of writing. It revealed that a large class of superior ninth-grade students can progress in composition as well as a small class.

5. The Becker study compared three methods of teaching, described as normal, bibliographic, and kinescope methods. The so-called normal method was found to have a slight superiority over the others.

The conclusion of the Braddock investigations is that of the more than a thousand items originally noted, only these five may be considered to yield valid answers from sound research.

In a review of the Braddock report in College English, Jean Hagstrom (1964) analyzes the research methods employed in these five researches, and concludes, "It is therefore extremely disheartening to say that 1) none of them strikes a layman as definitive or persuasive, and 2) there is very little promise that without rigorous antecedent thought, the 'scientific' method applied to composition will yield better results in the future than it has in the past." He urges antecedent thought on such basic problems as: How do you find a subject? How do you clothe it in appropriate style? How do you give it effective organization?

The Research Foundation of the National Council of Teachers of English planned a 1967 conference of specialists in a number of fields to attempt to identify the fundamentals of composition.

Current research in composition has taken an interesting and profitable direction in applying statistical procedures to rhetoric. Kellogg W. Hunt (1965a, 1965b) collected large samples of students' writing at grades 4, 8, and 12. These were subjected to structural analysis to attempt to describe the differences in sentence structure as students mature. As a tool of analysis Hunt developed what he called the "T-unit": a group of words minimal in length, each terminated grammatically between a capital letter and a period. This unit has the advantage of preserving all of the subordination achieved by a student, and all of his coordination between words and phrases and subordinate clauses.

His findings, very much condensed, are shown in Table 1. This table reveals that the most significant mark of maturity in writing is not sentence length in itself, but the increase of length of the T-units. When superior adults writing for Harpers and the Atlantic Monthly were analyzed by the same means, the length of their T-units increased by 40 percent over the twelfth graders', an increase equal to the gain of twelfth graders over fourth graders. Twelfth-grade students came close to the professionals in the number of clauses they use; the growth among professionals is in the increased number of words per clause. These results give a clear indication of the factors to be studied and taught in the improvement of sentence structure.

Francis Christensen (1963b) counted sentence openers to provide objective knowledge of how professionals begin sentences. High-school and college composition teachers have generally encouraged a wide variety of sentence openers, including prepositional phrases of all types, participial phrases, gerund phrases, absolute constructions, and sentence inversions. One writer says, "When a student opens a sentence with an infinitive phrase or a past participle, we immediately stamp him as more mature..."

Christensen analyzed twenty recent or contemporary writers - ten of narrative and ten of expository writing - to determine by what constructions they open sentences. The findings are in direct opposition to the traditional teaching: the professionals place something before the subject in only a fourth of their sentences (24.47 percent); that is, over 75 percent of their sentences begin with the subject. Of those opening by something other than the subject 23 percent of the total sentences open with adverbial modifiers, 1.17 percent use verbal groups, and the inverted structure appears in only one of every three hundred sentences (0.32 percent of the total sentences). The excessive use of verbal groupings as openers produces a style which Christensen calls "pretzel prose": that is, twisted sentences. He concludes: "The good teacher should not only base his teaching on the practice of professional writers, but should himself practice before he preaches."

In another study Christensen (1963a) develops what he calls "A Generative Rhetoric of the Sentence." He presents a scheme for the purpose of "buckling down to the hard work of making a difference in the student's understanding and manipulation of language." From his own analysis and that of others he draws four principles underlying the rhetoric of the sentence:

1. Composition is essentially a process of addition.
2. The cumulative sentence requires direction of modification or direction of movement.
3. Additions are made in levels of generality or levels of abstraction.
4. Density of texture is achieved by adding frequently and much to nouns, verbs, and main clauses.

TABLE I

MARK OF MATURITY IN WRITING

	Average length of clauses	Ratio of clauses per T-unit	Average length of T-units	Ratio of T-units per sentence	Average length of sentences
Grade 4	6.6 words 77 percent	1.30 77 percent	8.6 words 60 percent	1.6 137 percent	13.5 words 80 percent
Grade 8	8.1 words 94 percent	1.42 85 percent	11.5 words 80 percent	1.37 117 percent	15.9 words 94 percent
Grade 12	8.6 words 100 percent	1.68 100 percent	14.4 100 percent	1.17 100 percent	16.9 words 100 percent

Using visual analysis Christensen shows how sentences are built (or generated) by the process of coordination (addition of parallel elements), by subordination, (addition of levels of modification) or by a combination of coordination and subordination. In essence Christensen recommends and demonstrates the superiority of the cumulative sentence over the complex sentence. His technique of analysis can enrich the study of literature as well as develop mature writing.

Two years later Christensen extended his theory of the sentence to the study of the paragraph. Using a parallel title "A Generative Rhetoric of the Paragraph" (Christensen, 1965), he shows that the principles of analysis he developed for sentences apply with equal cogency to the paragraph; that is, paragraphs are created by addition, involving coordination of elements, subordination of elements, or combinations of the two. As well as addition, there must be also direction of movement; there will be levels of abstraction, and varying density of texture. In applying these principles he formulated them under nine headings, a few of which are:

1. The paragraph may be defined as a sequence of structurally related sentences.
2. The top sentence of the sequence is the topic sentence.
4. Simple sequences are of two sorts: coordinate and subordinate.
6. Some paragraphs have no top, no topic, sentence.

The nine headings are developed by exposition and illustration. The conclusion to be drawn is that the paragraph may be studied and taught as the sentence is studied and taught by Christensen.

A. L. Becker (1965) offers a different analysis of the paragraph under the title "A Tagmemic Approach to Paragraph Analysis." A Tagmeme is defined as the shortest grammatical unit; Becker calls them spots or slots in a system where substitution is possible. They are composites of form and function. Limiting his analysis to expository paragraphs, Becker finds two fundamental patterns. The first has three functional slots which he labels T (topic), R (restrictions), and I (illustration). These slots correspond to three levels of generality in paragraphs (cf. Christensen) and reflect a natural way of talking or writing about something.

The second major pattern has two slots labeled P (problem) and S (solution). The P may be a question or an assertion; the S is the answer or response. Where the S is extended, it tends to follow the TRI structure. The majority of expository paragraphs fall into these two basic patterns.

Variations in the patterns are caused by deletion, reordering, addition, and combination. The formal markers of internal tagmemic structure are graphic (indentation), lexical (equivalence classes and lexical transitions), grammatical (changes in the grammatical role of equivalence classes signal new slots or new paragraphs), and phonological (shifts in pitch, register, tempo, and volume in oral reading).

This analysis makes possible the partitioning of paragraphs in a consistent and predictable way.

The accurate and impartial evaluation of composition is one of the major problems of English teaching. Paul Diederich (1966) made a study of the factors that account for differences in the grades of qualified readers. Three hundred papers on two subjects were read by sixty distinguished readers representing six different fields: English, social studies, natural science, writers and editors, lawyers, and business executives. By comparison of the ratings of three readers who stood highest in a particular factor with three who stood lowest in the same factor, it was possible to isolate influential factors. In all, 11,018 comments on 3,557 papers were tabulated.

The factors revealed stood in this order of frequency: ideas, mechanics, organization, wording, style or "flavor." For experimentation in high schools, these were divided into two groups: I. General Merit (ideas, organization, wording, and flavor); II. Mechanics (usage, sentence structure, punctuation, capitals, abbreviations, spelling, hand-

TABLE 2

DISTRIBUTION OF RATINGS

	Non-Academic (N = 476)		
Grade	10 (145)	11 (167)	12
High	5 percent	8 percent	9 percent
Middle	34 percent	53 percent	63 percent
Low	61 percent	39 percent	28 percent

writing, neatness). A rating card was developed for the use of evaluators. A close following of the techniques developed by Diederich, with a year of practice for a faculty, should yield a reliability of the cumulative total for each student close to .90.

In the same article Diederich (1966) develops an elaborate but practical procedure for the rating of large numbers of papers. By a process of sorting and grouping of anonymous papers of combined grade groups, such as grades 10, 11, and 12, the process results in precentage figures for selected groups like Non-Academic and Academic with significant differences for High, Middle, and Low in each group, and a recognizable growth figure from grade 10 to grade 12. This extract (see Table 2) from a senior-high population will illustrate the distribution of ratings.

In this lower ability group the progress in the highest level is slight but observable: 4 percent. The progress in the middle level is striking: from 34 percent to 63 percent: the reduction in the low group is equally striking: from 61 percent to 28 percent. The figures from the Academic group show great advances in the high and middle level. This procedure demonstrates to teachers, administrators, and parents a positive progression of improvement not demonstrable by customary paper-grading processes.

RESEARCH IN LANGUAGE LEARNING

Ruth Strickland (1962) undertook to analyze the structure of children's oral language in the first through the sixth grades, to compare it with the structure of language found in the texts designed to teach children to read, and to discover, at selected grade levels, the influence of any ostensible differences in the quality of children's reading skill. A major hypothesis was that a study of children's speech might offer suggestions for the construction of better textbooks in reading.

In the analysis of structure, twenty-five consecutive sentences recorded on tapes were used as the language sample for each child. A special scheme was devised for the analysis of the language at two levels. Level One was divided into "slots" or stationary elements, and "moveables," or elements which can appear in different positions. Level Two was concerned with the types of subordinations or "satellites" used in the fixed slots and the moveables. This structure analysis is an important contribution of the of the study. Tabulations from this analysis were programmed to computers to relate the frequencies to the variables of age, sex, intelligence, socio-economic status, and parent education. It was found that many similar patterns were used with great frequency by children of every grade. It was also noted that children of all levels used a wide range of language patterns, and could elaborate and expand sentences by the use of moveables and subordinate elements.

In the comparison of children's language with that of textbooks in reading, four sets of widely used textbooks were analyzed for language patterns. The patterns used by children and by texts at relevant grade levels were studied. No scheme for control over sentence structure to parallel that over vocabulary seemed to exist. This raises the question: Can knowledge of the structures used by children be used to devise textbooks controlling the difficulty of reading?

A study of the relationship between children's use of language and the skill they demonstrate in learning to read in grade 6 revealed that children ranking high in the variable elements made more use of the common oral structural patterns than did children ranking low; the same children also made greater use of moveables and elements of subordination than did children ranking low. Among the conclusions to be drawn are these:

a. Children's command of language patterns is much broader than is generally realized.
b. There is almost no relationship between children's oral patterns and the patterns of structure in reading texts.
c. Children can be helped to recognize the patterns of expression and by this means be aided to advance in reading and in the learning of grammar and usage.

Walter Loban, of the University of California, began in 1952 a longitudinal study of children's growth in language which continued for eleven years. His report The Language of Elementary School Children (Loban, 1963) covers the first seven years, the elementary school level. From the Oakland, California schools 338 subjects were chosen at kindergarten level and were observed and tested in the following years. Items especially studied were: their vocabulary, their use of oral and written language, their proficiency in reading and listening, teachers' judgments of their skill with language, and background information on their health and homes.

Loban developed what he calls "the communication unit" in analyzing children's speech. It may be defined as a group of words which cannot be further divided without the loss of their essential meaning. Another term he developed is "the maze," a series of words or initial parts of words which do not add up to meaningful communication. Language samples recorded on disks and tapes provided the source materials. Two sub-groups, one exceptionally high and one exceptionally low in language ability, were identified.

A large number of specific findings have been summarized in the published report as follows:

During the first seven years of schooling, children speak more words in each succeeding year, produce more communication units, and increase the average number of words in those units.

Children rated as skillful in language reduce both their incidence of mazes and the number of words per maze.

Not pattern, but what is done to achieve flexibility within the pattern proves to be a measure of effectiveness and control of language at this level of language development.

Children who are rated as most proficient in language are also those who manifest the most sensitivity to the conventions of language.

Those high in general language ability are also high in reading ability. . . Writing ability is related to socioeconomic position . . . The highest correlation in the study is between vocabulary and intelligence . . . A low but positive relation exists between health and language proficiency . . . Competence in the spoken language appears to be basic for competence in reading and writing.

An important aspect of language learning in children is their ability to learn and use standard English. Loban in Problems in Oral English (1966) states that for most children a command of standard English is necessary, for "society exacts penalties of individuals who do not possess it." In continuing his longitudinal study of children's language, Loban here examines examples of nonstandard oral usage among four different groups of school children from kindergarten through grade 9. His chief findings have been summarized thus:

In ten years of schooling, pupils from homes in which class dialect is used make almost no improvement in using the verb to be appropriately or in standardized verb forms.

The practice, so common among weaker teachers, of drilling all pupils on the same skill is not supported by this research. Individual pupils, but not whole classes of pupils, will need help if they are to use the standard forms of irregular verbs.

As the complexity of sentence structure and total volume of spoken language both increase, there is a more than proportional probability of difficulty with . . . problems of clarity and precision, not problems of habitual usage.

Almost all the pupils whose parents speak informal standard English have little need of drill on usage.

Oral drill is more effective than workbook drills.

THE EMERGING GRAMMARS AND DIALECT STUDY

In the last fifteen years a fundamental tenet of English teaching, namely, that there is one grammar of English to be learned and taught, has been blasted by the development of at least two other systems of English grammar and the reasonable expectancy of others to follow. It is no longer accurate to speak of the grammar of English except in reference to the complete operation of the English language. We now have several grammars, each a system to describe how the English language works, or more specifically, how English makes sentences. The traditional system was developed principally in the eighteenth century by scholars whose grammatical disciplines were derived from Latin and Greek. Traditional grammar, therefore, was developed to make English conform to structural principles of Latin and Greek. Aside from certain classical biases, and lack of then undeveloped linguistic principles, it is a system of analysis of English of great value in the past and is by no means obsolete at present, unless viewed as the grammar of English.

In 1952 C. C. Fries introduced structural grammar. He was chief among a number of scholars working along parallel lines. The contribution of structural grammar is accurate description: it is concerned with the detailed elements of a language, its sounds, word units, inflections, and syntax. In the latter, analysis is conducted on the basis of immediate constituents as operating in modification, predication, complementation, and subordination. Structural grammar has greatly enriched knowledge about the elements of English and the manner in which they are combined to express meaning.

Noam Chomsky (1957) of MIT brought out in published form a new approach to grammar under the title Syntactic Structures. Somewhat modified by revisions, this newest of grammatical theories has great promise for the future. Generally called generative-transformational grammar, it is a system which attempts to state the "rules" which generate all English sentences and only English sentences. It assumes that we have in us a mechanism for using these laws which we gain from intuition. We are therefore able to generate sentences that have never been uttered before. The analysis begins with kernel sentences, simple declarative sentences without development. From these kernels other sentences are derived by means of rules for transformations, rules which change or develop kernel sentences containing phrase structures which are grammatical parts of a grammatical sentence. The goal of this new grammar is to state the rules that will make the production of sentences predictable.

Applications of structural and transformational grammar for classroom purposes are developed by Owen Thomas (1965) and Paul Roberts (1964). The former develops the transformational system in easy stages for the English teacher to understand and utilize the processes of generative-transformational grammar in the instruction of students. The latter presents a set of programmed lessons for self-instruction in the new grammar. N. S. Blount (Blount and Johnson, 1966) of Madison, Wisconsin, has created a similar programmed series of lessons adapted to the eighth grade.

The following curriculum question arises: "What happens to students who are taught the structure of English by generative-transformational processes?" Bateman and Zidonis (1966) offer a tentative answer

as the result of an experiment with ninth-and tenth-grade students. Using equivalent experimental and control groups, they taught the former by transformational techniques and the latter by traditional techniques. A valid sampling of free writing of each student was collected at the beginning and end of the instruction. The analysis of these writings by the Hunt (1965b) process revealed that the experimental group wrote better sentences of more sophisticated structure than did the control group. A parallel experiment is now under way at Manitowoc, Wisconsin, under the leadership of N. S. Blount of The University of Wisconsin, Madison.

A Dictionary of American Regional English is in progress, centered at The University of Wisconsin and directed by F. G. Cassidy of the Department of English (1966). Such a dictionary has been planned for over seventy-five years, and active collecting has been going on since 1889. Now under a 5-year contract with the U. S. Office of Education the work is progressing rapidly. Collections of data for six states, including Wisconsin are completed; six more are partially completed. Work of a similar nature is well under way in eleven more states. Many of the field workers live and work in "Word Wagons," specially equipped small buses containing the essentials of living quarters and apparatus for field collection.

Data are assembled, classified, and placed in a computer file at Madison. Collectors record speech; readers analyze regional literature and local newspapers. The effort is to secure complete coverage of the words and expressions used differently in various areas of the United States.

Curriculum applications of this study have yet to be worked out, but they will include 1) greater and more precise knowledge of American dialects; 2) a classification of usage levels and standards; and 3) new attitudes toward local and regional language variations.

WISCONSIN ENGLISH CURRICULUM ACTIVITY

By means of a Federal contract initiated in May, 1963, the State of Wisconsin has been served by the Wisconsin-English-Language-Arts Curriculum Project, jointly sponsored by The University of Wisconsin and the Department of Public Instruction. By means of two series of bulletins, conferences, conventions, and personal school visits, more than 5,000 elementary-school and high-school English teachers have been involved, many very actively. The project is, in fact, a statewide in-service training program in the English language arts. The aim is to encourage teachers to think, confer, discuss, and write about the curriculum in English. They have responded magnificantly. Over 3,000 bulletins go out each month; in the larger cities these are duplicated and distributed to additional numbers of teachers. By means of the reports of group leaders the project director is informed of the number of teachers who are meeting, the grade levels they represent, and their views on the issues that are raised.

In the summer of 1964 a selected group of teachers representing all school levels met to produce a curriculum in the teaching of literature – a cumulative growth pattern from the kindergarten through grade 12. After some months of revisions and editing, this was published in January of 1965 as Teaching Literature in Wisconsin, a paper-bound volume of 160 pages. Nearly 5,000 copies were distributed free in Wisconsin; it is estimated that the total distribution, including sales at cost, is about 17,000 copies.

By a similar process in the summer of 1965 a new group of selected teachers prepared the materials for the volume Teaching Speaking and Writing in Wisconsin (1966). Over 4,000 copies were distributed free; a large number have been sold at cost.

Again in the summer of 1966 two curriculum conferences were held, the first composed of ten selected college or university teachers who used their linguistic knowledge and experience in the teaching of English to plan the language curriculum. This was followed by a group of teachers representing elementary, junior high school, and senior high school levels which was to work out a continuum of language learning from kindergarten through grade 12. The result of this creative work and subsequent editing was issued in February of 1967 under the title Teaching the English Language in Wisconsin. Featured in this volume are:

1. A linguistic approach to grammar, utilizing simple applications of structural and transformational grammar to the language learning of children and youth.
2. A program that is sequential and non-repetitive.
3. A grammar that provides for more sophisticated sentence analysis at the time students need it.
4. An enrichment of the language program by the inclusion of word-study, semantics, and the history of the English language.

The three curriculum volumes listed above are issued in experimental editions. They were thoroughly revised in the light of experience in the classroom and reissued in the Spring of 1968 as one volume entitled English-Language-Arts in Wisconsin.

REFERENCES

Bateman, D., and Zidonis, F. K. The effect of a study of transformational grammar on the writing of ninth and tenth graders. Research Report No. 6. Champaign, Illinois: National Council of Teachers of English, 1966.

Becker, A. L. A tagmemic approach to paragraph analysis. College Composition and Communication, 1965, 16, 237-242.

Blount, N. S., and Johnson, Shelby L. Selected concepts in structural and transformational grammar. Twenty-two unpublished lessons, Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, 1966.

Braddock, R., Lloyd-Jones, R., and Schoer, L. Research in written composition. Champaign,

Illinois: National Council of Teachers of English, 1963.

Cassidy, F. G. Dictionary of American regional English. Unpublished progress report, 1966.

Chomsky, N. Syntactic structures. The Hague, Netherlands: Mouton, 1957.

Christensen, F. A generative rhetoric of the sentence. College Composition and Communication, 1963, 14, 155-161. (a)

Christensen, F. Notes toward a new rhetoric. College English, 1963, 25, 7-18. (b)

Christensen, F. A generative rhetoric of the paragraph. College Composition and Communication, 1965, 16, 144-156.

College Entrance Examination Board. Freedom and discipline in English. Princeton, New Jersey: CEEB, 1965.

Diederich, P. B. How to measure growth in writing ability. English Journal, 1966, 55, 435-449.

Fries, C. C. Structural grammar. New York: Harcourt, Brace and World, 1952.

Grommen, A. (Editor) The education of teachers of English. New York: Appleton-Century-Crofts, 1963.

Hagstrom, J. H. Review of National Council of Teachers of English, Research in written composition. College English, 1964, 26, 53-56.

Hunt, K. W. A synopsis of clause-to-sentence length factors. English Journal, 1965, 54, 300-309. (a)

Hunt, K. W. Grammatical structures written at three grade levels. Research Report No. 3. Champaign, Illinois: National Council of Teachers of English, 1965. (b)

Loban, W. The language of elementary school children. Research Report No. 1. Champaign, Illinois: National Council of Teachers of English, 1963.

Loban, W. Problems in oral English. Research Report No. 5. Champaign, Illinois: National Council of Teachers of English, 1966.

National Council of Teachers of English. The national interest and the teaching of English: A report on the status of the profession. Champaign, Illinois: NCTE, 1961.

National Council of Teachers of English. The national interest and the continuing education of teachers of English. Champaign, Illinois: NCTE, 1963.

Roberts, P. English syntax. New York: Harcourt, Brace, and World, 1964.

Squire, J. R. The responses of adolescents while reading four short stories. Research Report No. 2. Champaign, Illinois: National Council of Teachers of English, 1963.

Strickland, Ruth G. The language of elementary school children. Bulletin of the School of Education, Indiana University, 1962, 38, No. 4.

Thomas, O. Transformation grammar and the teacher of English. New York: Holt, Rinehart, and Winston, 1965.

Wilson, J. R. Responses of college freshmen to three novels. Research Report No. 7. Champaign, Illinois: National Council of Teachers of English, 1966.

Wisconsin English Language Arts Curriculum Project. Teaching literature in Wisconsin. Madison, Wisconsin: Department of Public Instruction, 1965.

Wisconsin English Language Arts Curriculum Project. Teaching speaking and writing in Wisconsin. Madison, Wisconsin: Department of Public Instruction, 1966.

Wisconsin English Language Arts Curriculum Project. Teaching the English Language in Wisconsin. Madison, Wisconsin: Department of Public Instruction, 1967.

Research in Reading[1]

Wayne Otto and Thomas C. Barrett
The University of Wisconsin
Madison, Wisconsin

Theodore L. Harris
University of Puget Sound
Tacoma, Washington

THE COMMITTEE for Research in Basic Skills has spent a number of years investigating problems of teaching and learning fundamental skills in the elementary grades. Generally, the Committee has chosen to be extensive in its approach, investigating many dimensions of a given problem. As a result, there have been a series of studies within broad problem areas.

The present investigation grew out of a recurring concern for pupils' development of flexibility in performance in the basic skill areas. In a project entitled "The Perception of Symbols in Skill Learning," Herrick, Harris, and Rarick (1961) established the importance of the development of variability rather than rigidity in motor-speed sets of handwriting. The authors pointed out that their findings might well have implications for other areas of skills development. The question of flexibility arose again in two studies (Harris and Rarick, 1957; Herrick and Otto, 1961) of the handwriting act. One implication of these studies was that efficiency in a skill such as handwriting does not seem to be associated with the development of a fixed level of pressure but rather with the ability to control pressure within given limits. Another study by Otto (1961) indicated that among good readers in the elementary grades responsiveness to auditory, visual, kinesthetic, and combined presentations of stimuli varies according to grade level and type of reinforcement. The variance, however, was not always in the direction expected, which suggests that one effect of the instructional program may be the development of perceptual sets in reading.

These studies led the Committee to the U. S. Office of Education Project No. 1755, "The Experimental Development of Variability in Rate of Reading in the Intermediate Grades" (Harris, Herrick, Macdonald, and Rarick, 1965). The results of the study showed that fourth-grade good readers who received a 2-week training program developed substantial ability to vary their speed when reading for different purposes. In the training program the children were given a) a brief introduction to three types of reading purposes – reading for the main idea, reading to find a specific fact, and reading to follow the sequence of ideas, b) short testing and training passages carefully constructed as to unity, coherence, emphasis, and appropriate content and difficulty level, c) separate selections for narrative and expository style, and d) testing to a criterion of full comprehension of each specified purpose. Specifically, the children were shown to a) possess a hitherto unreported ability to adjust significantly their reading speeds very quickly to their purposes for reading; b) alter significantly their pattern of reading speed relative to reading purposes and to maintain a significant spread in speed-purpose variability while increasing their reading speeds for each purpose approximately 40 percent, with further gains upon retention tests two weeks after training; c) adjust their reading speeds to reading purposes significantly different according to passage style and to a certain extent according to sex; d) and display the greatest improvement in reading for the main idea and specific fact. The results of Project 1755 also showed that a) children of comparable reading ability in the fifth and sixth grades have no more variability or spread in their reading speeds

according to reading purposes than do fourth graders; b) children, by their own daily ratings, greatly enjoyed the 2-week training period which was largely self-instructional in nature; c) during the short training period, children became significantly more realistic in predicting their reading speeds when asked to read for sequence of ideas, but not for main idea or specific fact; and d) no untoward evidences of tension were found to be associated with reading the test passages or answering questions on them.

In keeping with its past policies, the Committee wanted to relate directly its research tasks to specific aspects of school practice or to provide the basis upon which its findings could be generalized to school programs. The experimenters felt that before schools could be expected to implement the findings of Project 1755 in their programs, they should have empirical evidence that the results of training given under the aforesaid experimental conditions would indeed transfer to more typical reading tasks faced by intermediate-grade children. The present study is, therefore, an extension of Project 1755 specifically designed to investigate the possibilities of such transfer effects.

The first part of the present study replicated the training and testing of Project 1755. The only major revision in the first part of the study was in the substitution of a second active learning group for a passive control group. Since the preceding study gave evidence that trained groups achieved greater variability than controls, it was felt to be more pertinent to test for significant differences between types of training. Therefore, half of the subjects were trained with passages containing an explicitly stated main idea and the other half worked with passages in which a main idea was implicit but not stated. Classroom materials may or may not contain explicit main idea statements so the intent was to determine whether training with either type would produce better results.

Three major hypotheses were stated:

1. Intermediate-grade children trained with passages containing either explicitly stated or implicit main ideas will differ in rate variability when reading carefully structured expository or narrative materials for three different purposes.
2. Intermediate-grade children trained with passages containing either explicitly stated or implicit main ideas will transfer their variability of reading rate to more typical materials.
3. Intermediate-grade children trained with passages containing either explicitly stated or implicit main ideas will retain their variability of reading rate with typical materials after a period of one month.

Answers to the following questions were sought:

1. Do children differ in their total reading rate when reading either narrative or expository material for three different purposes?
2. Do boys and girls in grades 4, 5, and 6 trained with passages containing specifically stated main ideas and children trained with passages in which main ideas were implicit, but not stated, differ in their rate variability when reading either expository or narrative materials for three different purposes on a training-like task immediately after training?
3. Do boys and girls in grades 4, 5, and 6 trained with passages containing either explicitly stated or implicit main ideas differ in their rate variability when reading either expository or narrative materials for three different purposes on a transfer task immediately after training?
4. Do boys and girls in grades 4, 5, and 6 trained with passages containing either explicitly stated or implicit main ideas differ in their rate variability when reading either expository or narrative materials for three different purposes on a a) training-like task immediately after training; b) transfer task immediately after training; c) transfer task one month after training?
5. Do children vary their rate when reading either expository or narrative materials for three different purposes – specific fact, sequence, or main idea on a a) training-like task immediately after training; b) transfer task immediately after training; c) transfer task one month after training?

METHOD

SUBJECTS

In the final analyses, seventy-two children from grades 4, 5, and 6 of a middle-class elementary school in Madison, Wisconsin, served as subjects. There were twenty-four subjects at each grade level, twelve boys and twelve girls. In an attempt to exclude the extremes of reading ability, subjects were chosen who had reading test scores that placed them between the fortieth and ninetieth percentiles for their grades. The STEP Reading Test was used in grade 4, the Reading Comprehension subtest of the California Achievement Test was used in grade 5, and the Iowa Reading Test was used in grade 6. School personnel corroborated the reading test scores. Table 1 shows the percentile range by grade and sex.

Within sex and grade level, equal numbers of subjects were randomly assigned to the two treatment groups. To guard against attrition due to absence, inability to cope with the task, and examiner errors, extra subjects, one boy and one girl from each grade level, were included in each treatment group. Therefore, the treatment groups comprised forty-two children, of whom six were not included in the analysis. The extra subjects remaining after attrition were randomly dropped.

EXPERIMENTAL TREATMENTS[1]

The distinction between groups was in 1) the nature of orientation to reading for the main idea and 2) the structure of materials read during the orientation and training periods. During these periods, Group I worked with materials[2] in which a main idea was explicitly stated; the task in reading for the main idea was to identify the main idea from a list of sentences taken from the passages read.

TABLE 1

PERCENTILE RANGE OF SUBJECTS' READING TEST SCORES BY GRADE AND SEX

	Grade					
	Fourth		Fifth		Sixth	
Sex	Highest Percentile	Lowest Percentile	Highest Percentile	Lowest Percentile	Highest Percentile	Lowest Percentile
Boys	88	48	90	62	87	40
Girls	90	50	90	54	89	57
Total	90	48	90	54	89	40

Group II worked with the materials from which the stated main idea had been deleted; the task in reading for the main idea involved formulating a main idea and then selecting the main idea sentence from a list identical to the one used by Group I. Both groups worked with passages of identical length and essentially the same substantive subject matter.

MATERIALS

Materials for orientation to the task, for variability training, and for immediate posttraining evaluation were taken, with but minor adaptations explained below, from materials constructed for Project 1755. Longer materials, more representative of the type encountered by elementary pupils, were constructed for use in the attempts to assess the transfer effects of the variability training. The development of materials was a major focus in both the present project and the one that preceded it; thus, the materials are described in considerable detail.

ORIENTATION MATERIALS

The orientation materials comprised three sets of passages, one set for each of the three purposes for reading (see description of the training program). Each passage was printed separately on charts which could be read from the location of any pupil in the testing room. A list of five names with corresponding telephone numbers was used in finding a specific fact. Reading for sequence involved one basic paragraph with three different orderings of its five sentences. For Group I orientation to reading for the main idea also employed a basic paragraph which was rewritten three times so that the topic sentence appeared in different parts of the passage. The materials were taken from those developed for Project 1755.

Because all of the Project 1755 materials included explicitly stated main ideas, the following adaptations were required for use with Group II. First, the sentence stating the main idea was deleted from the passage used to orient the subjects to reading for the main idea passage and the remaining sentences were expanded slightly to preserve a uniform length. Second, an additional paragraph was used with Group II during its orientation to the purpose of reading for the main idea. This paragraph was similar to the first one in topic and construction, but it was used to provide further clarification of the task and practice in arriving independently at a conception of the main idea. A pilot study showed that the added practice was needed.

TRAINING MATERIALS

The training materials developed in Project 1755 and used in the present study consisted of three parts; a set of directions, the passages to be read, and comprehension checks. Brief descriptions of each of the three parts of the training materials follow.

Directions

The primary purpose of the directions was to establish a definite purpose for reading either to find the Main Idea, The Sequence of Events, or a Specific Fact. Thus, a specific question was asked for each selection to be read for a Specific Fact, and when the purpose was to read for Sequence or Main Idea, the subject was told to read to find the Main Idea or to remember the order of events. The directions were designed to make the training self-instructional.

Passages

The passages were written with the following restrictions and considerations in mind. The content of the training materials was from the areas of social studies, science, or fiction because Shores' (1960) research indicates that reading speed is affected by different kinds of content.

The passages were written in narrative and expository styles because a pilot study in Project 1755 showed that expository materials were read faster than narrative, yet pupils commonly encounter both styles.

Passages were given either an inductive structure (i.e., the topic sentence came near the end of the

selection) or a deductive structure (i.e., the topic sentence came near the beginning of the selection). The decision to consider structure of the material was based upon results of pilot testing in Project 1755 which indicated that children perform better on paragraphs organized deductively than inductively.

Passages were limited to six to eight sentences with total word count ranging from 66 to 112 words since the earlier pilot study showed that fourth-grade students had much difficulty with longer selections before training.

In an attempt to keep the passages within the fourth-grade readability level according to the Dale-Chall formula, the vocabulary was limited to grade 4 word lists, sentence length was limited to a seven to fourteen words range, and typical subject-verb word order was maintained within the sentences. Each passage was allowed a maximum of two examples of the syntactical complication which follows: relative and subordinate clauses, participial phrases, infinitive phrases separated from the main verb, and inverted subject-verb word order. Five teachers with experience in the intermediate grades judged the difficulty of the materials to be appropriate. As an added check, the directions for each passage encouraged the subjects to ask for vocabulary help as needed.

Each passage has a single main idea. Group I practiced with passages in which a main idea was explicitly stated; Group II practiced with passages in which the same main idea was implied but not stated. In the passages where the main idea was not stated, adjectives, adverbs, and other words which did not change the meaning of the paragraphs were inserted in the remaining sentences to yield a constant paragraph length. The nature of the daily training materials is shown in Table 2.

Comprehension Tests

The position was taken that the reader should have access to all the passage units when selecting the appropriate main idea or determining the proper order of ideas. Thus, for the main idea and sequence tasks the comprehension tests contained a condensed or abridged restatement of each of the six to eight sentences in the passage itself and the subject either identified the main idea statement or sequentially ordered the statements given. For the specific fact tasks, a phrase containing a specific fact was taken from each sentence and the subject identified the correct one.

POST-TRAINING MATERIALS

The directions and comprehension tests used with the post-training materials were of the same type as the training materials proper and the passages were likewise adapted from paragraphs constructed for Project 1755. Six of the passages were written without a stated main idea and six were written with an explicit main idea. Each passage was sixty-six words long, including five sentences if the main idea was not stated or six sentences if it was stated.

TRANSFER MATERIALS

The central question of the study is whether variability in reading speed developed by training with

TABLE 2

NATURE OF THE DAILY TRAINING PROGRAM MATERIALS

Number	Selection
4	Main Idea Selections
1	Sequence Selection
1	Specific Fact Selection

carefully constructed, short model paragraphs will transfer to reading the longer, less tightly structured passages typically found in classroom material. Thus, the decision was to use the types of material actually employed in classroom instruction in the transfer task. With this decision, there was the problem of learning more about the nature and length of "typical" paragraphs in elementary level materials. A preliminary survey of tradebooks and textbooks, especially b... reading series, revealed a lack of consistency within intermediate level materials and even within individual publications.

It had at first been assumed that construction of the transfer materials would involve the use of a certain number of paragraphs, but the survey revealed that there was a surprising lack of correspondence between paragraphing and topical units. Within a single paragraph, two or three topics might be discussed; yet at another place in the same material several paragraphs might be used to cover one topic. Since one of the purposes for reading in the study involved the identification or formulation of a main idea, the experimenters decided to select passages that contained only one main idea, regardless of the total number of paragraphs involved. The survey showed also that there was little consistency among the materials with regard to the number of words used in the coverage of a single main idea. Because the experimental design of the study called for a uniform length in the transfer test materials, the experimenters arbitrarily set the length at 198 words, three times the number of words in the post-training test materials.

With consideration for all the factors involved in the looseness or tightness of the structure of materials at the intermediate grade levels, the experimenters decided that selections from the intermediate level Reader's Digest Skill Builders were reasonably representative of materials used by children in the fourth, fifth, and sixth grades.

Passages were chosen from the Skill Builders in view of the following: each passage had to have a single, complete main idea; each passage had to be appropriately written for one of the three purposes, (i.e., if the purpose were to read for sequence, the passage must obviously contain a clearly recognizable sequential development); and all of the passages chosen had to be generally interesting to intermediate-grade children but not specifically familiar to them. The Skill Builder selections were modified

only when necessary in terms of the addition or deletion of a stated main idea, and/or the addition or deletion of unimportant words in order to keep the length constant. (The authors are indebted to the editors of Reader's Digest for permission to use their materials and for their examination and approval of all adaptations made in the materials.) *Half of the selections chosen were written in a primarily narrative style and half were expository in style.*

Equivalent sets of materials were necessary for the immediate and delayed transfer testing. "Pairs" of materials were developed by drawing two selections from the same Skill Builder article. Since these articles were very short, it was felt that vocabulary, style, and other factors affecting readability were likely to be reasonably equivalent. The selections in each pair were assigned randomly to the immediate or delayed transfer testing materials.

Table 3 is a breakdown of the total number of selections used for the transfer study.

TABLE 3

NATURE OF THE SELECTIONS USED IN THE IMMEDIATE AND DELAYED TRANSFER SESSIONS

Selections for Immediate Transfer		Selections for Delayed Transfer	
Stated Main Idea			
2*	Main Idea	2	Main Idea
2	Sequence	2	Sequence
2	Specific Fact	2	Specific Fact
Unstated Main Idea			
2	Main Idea	2	Main Idea
2	Sequence	2	Sequence
2	Specific Fact	2	Specific Fact

*There was one narrative and one expository selection in each pair.

TRAINING AND TESTING PROCEDURE

In keeping with the framework established by Project 1755, the research program was divided into two basic parts, training and testing, which were then subdivided with time designated for each part as shown in Table 4.

Training Program

Training Examiners. A week before the training program began, the adult assistants met with the graduate research staff for orientation to methods and procedures to be used. A brief history of the research project was given, and each assistant received a copy of the complete set of directions for each day of training. Objectivity in working with the subjects was emphasized.

Training Subjects. Days 1-3 of the training program served as an Orientation Period to prepare subjects to read for the three different purposes and to familiarize them with the materials and procedures. The subjects read three practice passages each day. The orientation sequence is shown in Table 5.

The difference between groups in the Orientation Period was in presentation of reading for the main idea. Group I practiced with stated main ideas in their paragraphs while Group II practiced with paragraphs with implicit main ideas. This distinction was maintained throughout the training program.

On days 4-10 the subjects practiced for 30-45 minute periods each day, during which they read six selections daily. The selections were randomly ordered for the total group of eighty-four subjects every day.

In addition to reading passages, subjects also practiced predicting their reading speed for each selection. This was followed by recording actual speed on their individual graphs and on Time Record Sheets. It was felt that this would help each subject evaluate his own progress; the graph would enable him to visualize the differences in reading rate for each of the three purposes for reading.

On each of the seven practice days, this procedure was followed:

1. Each subject received his set of self-instructional materials and read the directions to establish a purpose for reading.
2. Each subject predicted his reading speed on his graph.
3. Each subject timed his reading of the selection with a stop watch and recorded actual reading time on Time Record Sheets.
4. An answer sheet accompanying the selection was then completed.
5. Subjects were expected to read to full comprehension or 100 percent accuracy in responding to the comprehension test items. Due to the nature of the tests, five trials was the maximum. The assistants checked individual responses after each trial.
6. Raw time scores were translated into words per minute scores by the subjects, who were supplied with conversion tables.
7. Subjects then plotted reading speed on their graphs to show a comparison of predicted and actual reading speed. The graphing was intended to serve as a motivational device.

ADMINISTRATION OF THE TRAINING PROGRAM

To obtain manageable groups for the training sessions, Groups I and II were each divided into two training groups of twenty-one subjects. Three groups met separately in the testing room during the morning, and one group met in the afternoon.

Four research assistants conducted the Training Program. Since each had been involved in working on various phases of the training materials and training procedure, it was felt that they would be highly consistent in their administration of the four training groups. Four or five adult assistants were

TABLE 4

SEQUENCE OF TRAINING AND TESTING

Training Program		Testing Program		
Orientation Period	Training Period	Post-Training Test Period	Immediate Transfer Test Period	Delayed Transfer Test Period
Days 1-3	Days 4-10	Day 11	Days 12-13	Days 14-15 (one month following Days 12-13

TABLE 5

ORIENTATION SEQUENCE

Day 1	Day 2	Day 3
Introduction to (1) three purposes for reading (2) use of stop watches (3) use of Time Record Sheets	Introduction to (1) Conversion tables (2) Graphs	Introduction to (1) predicting reading speed

available at each session to supervise the subjects' work. The resulting ratio of one adult for every four or five subjects insured adequate supervision of the work and immediate feedback on the comprehension tests.

The following administrative details were patterned after Project 1755:

1. Training groups were no larger than twenty-one subjects.
2. A large, well-ventilated, well-lighted, quiet room was used.
3. Individual work space was available for each pupil.
4. All necessary supplies were provided by the researchers.
5. One adult was assigned to work with every four or five subjects.
6. Instructors and assistants were trained in the testing procedures and to anticipate possible problems.
7. Detailed printed instructions were provided for instructors and assistants.

TESTING PROGRAM

The procedure followed during the three phases of the Testing Program (see Table 4) deviated from the Training Program in these ways: (1) For the Post-Training Test, day 11, 12 randomly ordered, sixty-six word selections were read instead of the usual six selections of varying length. Four randomly ordered selections were read for each of the three purposes. The total reading time in seconds for each purpose was recorded. The pupils did not use their graphs. (2) The Immediate and Delayed Transfer selections were 168 words in length. They are described in detail in the materials section of this paper. All subjects read twelve randomly ordered selections, six with and six without an explicitly stated main idea, for Immediate Transfer on days 12 and 13. One month later, on days 14 and 15, they read twelve selections comprising the second half of the set of transfer materials. During the two transfer testing periods, the assistants recorded the subjects' total reading time in seconds when reading for each purpose. No graphs or conversion tables were used.

The assumption was that in order to assess the effects of the stated and implied main idea treatment conditions it was necessary to test the subjects on materials with and without explicitly stated main ideas. A reasonable expectation would be that subjects trained with stated main ideas would do best on test materials which contained stated main ideas and those trained with materials having implicit main ideas would do best on test materials with implicit main ideas. To assess the overall effect of the treatment conditions, the experimenters were concerned with the best performance on both types

of materials. In order to get a score which indicated this total performance, regardless of type of material, the mean score of the subjects' performance on narrative and expository material was found, thus reducing the number of scores for each subject to six.

DESIGN AND ANALYSIS

In order to determine the variability of reading rate for the three different purposes – specific fact, sequence, and main idea – a variability score was calculated for each subject. The style of the material was considered in that two rate variability scores were calculated for each subject – one for narrative and one for expository materials. The variability scores were calculated according to the standard variance formula:

$$S^2 = \sum_{i=1}^{3} \frac{(Xi - \bar{X})^2}{n - 1} .$$

Xi was the sum of individual scores where i varied across the three purposes; $\bar{X}$ was the mean of the three purpose scores, and n = 3. Thus, the variance scores took into account the subjects' reading rate for one purpose in relation to all purposes (mean rate). There were three different designs and consequently three different types of analysis.

One was a subject × purpose one way analysis of variance design. The dependent variable for the analysis was the proportion of each individual's reading rate for each single purpose to his total reading rate for all three purposes.

Due to the fact that each subject read for three purposes, a repeated measures situation existed. Greenhouse and Geisser (1959) have pointed out that for the usually computed F ratios of mean squares in the regular analysis of variance to be exactly distributed as the F distribution, it is necessary that the columns or tests in addition to being normally distributed, having equal variances, should be mutually independent. Since the measures were made on the same subjects, the independence assumption could not be met.

Although one approach to this problem would be to use a multivariate analysis of variance design, a simpler approach is to follow the procedure of adjusting the degrees of freedom which results in conservative F ratios which approximate the F distribution when the independence assumption is not met (Greenhouse and Geisser, 1959). The adjusted degrees of freedom are in this case 1 and 71 respectively. Six analyses were also done using this design and procedure.

The second was a 3 × 2 × 2 completely crossed analysis of variance design with three grades, two sexes, and two training methods and with six replicates per cell. The model was assumed to be fixed. The reading variability scores were transformed by using the natural log to base e. Thus, the scores (y) used in the analysis were:

$$y = \text{nat} \log_e \sum_{i=1}^{3} \frac{(Xi - \bar{X})^2}{n - 1} .$$

After the scores were transformed, an analysis of variance was done on the means of the variances according to the procedure outlined by Scheffé (1960) for the comparison of variances. Six analyses were done according to these design and analysis procedures.

The third design was a 3 × 2 × 2 × 3 analysis of variance with three grades, two sexes, two training procedures, and three repeated measurements or tests – a training test immediately after training, a transfer task immediately after training, and a delayed transfer task one month after training. The dependent variable for the analyses using this design was rate variability. Two analyses were done using this procedure. Conservative F tests using the Greenhouse and Geisser (1959) procedure were also used in this analysis. The adjusted degrees of freedom were 1 and 69 respectively. For all F ratios found to be significant post hoc comparisons among the means were done following the procedures of Scheffé (1953).

RESULTS AND DISCUSSION

The subjects consistently varied their rate when reading either narrative or expository material for three different purposes. The analyses summarized in Table 6 also show that there was a difference in rate variability on the immediate post training test, the transfer task, and the transfer retention task.

Mean proportions of reading time are shown in Table 7. The trend there is clear and, with but one exception, consistent: the children read most rapidly to find specific facts; they read less rapidly for main ideas, and they read most slowly for sequence. The Scheffé post hoc test for multiple comparisons reveals significant differences between all means except the Post-Narrative, Main Idea, and Specific Fact. This is in line with the logical expectation that, with training, children would be able to adapt their rate of reading according to purpose; that they would be best able to pick out specific facts very quickly; and that the task of reading to arrange facts in sequence would be most time consuming because the subjects would need to deal simultaneously with several facts.

Materials were written in narrative and expository styles because there was evidence that style affects rate. When the means in Table 7 were collapsed across style, there was virtually no difference between the narrative and expository means (Expository = .340; Narrative = .333). Yet, a comparison that takes purpose into account, as shown in Table 8, reveals that style interacts with purpose; expository materials are read most rapidly for the main idea, but narrative materials are read most rapidly for sequence and for specific fact. By way of post hoc interpretation, it would appear that the expository arrangement makes for more clearly identifiable main ideas. Likewise, the narrative arrangement of materials would tend to have a sequential arrangement built in and, therefore, be more easily identifiable. An explanation of the more rapid reading of narrative materials for specific fact is more elusive; but it might be argued that a narrative-sequential arrangement lends clues that make for rapid location of facts.

TABLE 6

ONE WAY ANALYSES OF VARIANCES FOR TOTAL READING RATE FOR THREE DIFFERENT PURPOSES WITH EXPOSITORY AND NARRATIVE MATERIALS ON POST, IMMEDIATE, AND DELAYED TESTS

Description	df	MS	F
Purposes on Expository Posttest	2 (1)	.40	40*
Error	213 (71)	.01	
Purposes on Narrative Posttest	2 (1)	.72	72*
Error	213 (71)	.01	
Purposes on Expository Transfer Test	2 (1)	.97	97*
Error	213 (71)	.01	
Purposes on Narrative Transfer Test	2 (1)	.92	92*
Error	213 (71)	.01	
Purposes on Expository Retention Test	2 (1)	.84	84*
Error	213 (71)	.01	
Purposes on Narrative Retention Test	2 (1)	.36	36*
Error	213 (71)	.01	

* $p < .01$

TABLE 7

MEAN PROPORTION OF READING TIME FOR TOTAL COMPREHENSION BY STYLE OF MATERIAL, TIME OF TEST AND PURPOSE

	Purpose		
Test and Style	Main Idea	Sequence	Fact
Post-Expository	.331	.409	.261
Post-Narrative	.389	.393	.218
Immediate Transfer-Expository	.359	.435	.207
Immediate Transfer-Narrative	.348	.438	.214
Delayed Transfer-Expository	.325	.446	.229
Delayed Transfer-Narrative	.382	.366	.252

TABLE 8

MEAN PROPORTION OF READING TIME BY STYLE AND PURPOSE

	Purpose		
Style of Material	Main Idea	Sequence	Fact
Expository	.338	.430	.252
Narrative	.373	.399	.228

TABLE 9

ANALYSIS OF VARIANCE OF GRADE, SEX, AND TREATMENT FOR EXPOSITORY MATERIALS AND POST PERIOD READING VARIABILITY FOR THREE PURPOSES

Source of Variation	df	MS	F
Grade	2	5.98	2.89
Sex	1	1.00	<1
Treatment	1	.55	<1
G × S	2	2.99	1.46
G × T	2	4.30	2.10
S × T	1	1.14	<1
G × S × T	2	.84	<1
Error	60	2.05	

TABLE 10

ANALYSIS OF VARIANCE OF GRADE, SEX, AND TREATMENT FOR NARRATIVE MATERIALS AND POST PERIOD READING VARIABILITY FOR THREE PURPOSES

Source of Variation	df	MS	F
Grade	2	3.76	1.52
Sex	1	2.07	<1
Treatment	1	4.11	1.66
G × S	2	.64	<1
G × T	2	.23	<1
S × T	1	1.17	<1
G × S × T	2	1.24	<1
Error	60	2.48	

The type of materials used during the training sessions did not make a difference in the children's subsequent variability rate when reading for three different purposes (see Tables 9-14). Children trained with passages containing explicitly stated main ideas did not differ in rate variability from children trained with materials with implicit but not specifically stated main ideas. The types of training materials are subsequently referred to as stated and unstated, respectively. The implication is that materials varied on this aspect do not affect children's ability to vary their reading rate. Thus, either type of training materials should produce similar results if reading variability across purposes is the objective of the training.

The use of different types of training materials did not have any effect on the transfer task. When children were given an extended passage which was more closely related to the actual school reading task, children trained with stated materials did not differ in their rate variability from children trained with unstated materials.

The same was true for the transfer retention task. When children were tested for retention on the transfer task, no differences in rate variability were found between the two groups trained with different materials.

However, one might speculate that if the approach to instruction had been varied for each type of materials, the results might have been different. For example, if the children who worked with unstated materials had been trained explicitly in devising their own main idea statements, they might have done better than the other group when all subjects were tested on both stated and unstated materials. Such speculation seems to suggest a possible next step for research that focuses upon training programs for the development of rate variability.

Whether the children were reading expository or narrative materials, the results were the same. Consistently on the post, immediate transfer, and delayed transfer tests no differences were found between the group trained with stated main idea materials and the group trained with unstated main idea materials.

TABLE 11

ANALYSIS OF VARIANCE OF GRADE, SEX, AND TREATMENT FOR EXPOSITORY MATERIALS AND IMMEDIATE PERIOD READING VARIABILITY FOR THREE PURPOSES

Source of Variation	df	MS	F
Grade	2	4.94	2.66
Sex	1	4.22	2.27
Treatment	1	.34	<1
G × S	2	2.35	1.26
G × T	2	2.40	1.29
S × T	1	4.43	2.38
G × S × T	2	2.40	1.29
Error	60	1.86	

TABLE 12

ANALYSIS OF VARIANCE OF GRADE, SEX, AND TREATMENT FOR NARRATIVE MATERIALS AND IMMEDIATE PERIOD READING VARIABILITY FOR THREE PURPOSES

Source of Variation	df	MS	F
Grade	2	.71	<1
Sex	1	5.33	1.60
Treatment	1	.00	<1
G × S	2	1.51	<1
G × T	2	5.10	1.53
S × T	1	.32	<1
G × S × T	2	.03	<1
Error	60	3.33	

Sex did not make a difference in the rate variability of children trained with either of the two types of materials. This result was consistent. Boys did not differ from girls in their ability to vary their rate from purpose to purpose immediately after training, on the transfer task or on the transfer retention task.

The same results held across grade level. There was no difference among fourth, fifth, and sixth graders in reading rate variability from purpose to purpose on each separate testing occasion. However, the trends indicated by the means (Table 15) are consistent. Fourth graders consistently showed more variability of reading rate than did fifth graders, and fifth graders consistently showed more variability of reading rate than did the sixth graders. This could be due to the younger children's more flexible mental set, which could make them much more susceptible to training.

The absence of grade by treatment interaction indicates that the two different training procedures did not have different effects across grade level. This might not have been so if materials had been prepared for each specific grade level. Further investigation that focuses upon possible developmental trends might be worthwhile.

The total reading variability across occasions (tests) on expository materials was found to differ by grade (Table 16). No grade differences were apparent, however, on the total reading variability across occasions on narrative materials (Table 17). Mean rate variability by grade and type of material is shown in Table 18. Scheffé *post hoc* tests for multiple comparisons revealed that the only significant difference is between the mean scores of fourth and sixth graders when reading expository material. This supports the trend found with individual tests. By way of interpretation, the suggestion is that older pupils may have a fixed set for approaching the reading task; whereas, younger pupils may be more readily susceptible to rate variability training. If this is so, the implication is that such training ought to be given early in the intermediate grades.

TABLE 13

ANALYSIS OF VARIANCE OF GRADE, SEX, AND TREATMENT FOR EXPOSITORY MATERIALS AND DELAYED PERIOD READING VARIABILITY FOR THREE PURPOSES

Source of Variation	df	MS	F
Grade	2	2.97	1.40
Sex	1	.16	<1
Treatment	1	.27	<1
G × S	2	.23	<1
G × T	2	.31	<1
S × T	1	.71	<1
G × S × T	2	1.93	<1
Error	60	2.12	

TABLE 14

ANALYSIS OF VARIANCE OF GRADE, SEX, AND TREATMENT FOR NARRATIVE MATERIALS AND DELAYED PERIOD READING VARIABILITY FOR THREE PURPOSES

Source of Variation	df	MS	F
Grade	2	5.80	2.10
Sex	1	1.78	<1
Treatment	1	.15	<1
G × S	2	2.95	1.07
G × T	2	2.45	<1
S × T	1	7.36	2.67
G × S × T	2	2.44	<1
Error	60	2.76	

TABLE 15

MEANS FOR EACH GRADE ON EXPOSITORY AND NARRATIVE MATERIALS ON THREE DIFFERENT TEST OCCASIONS

Type	Occasion	Grade 4	Grade 5	Grade 6
	Post	3.07	2.66	2.08
Expository	Immediate	6.29	6.20	5.46
	Delayed	6.40	5.89	5.72
	Post	3.45	3.35	2.72
Narrative	Immediate	5.66	5.92	5.59
	Delayed	5.80	5.61	4.87

TABLE 16

ANALYSIS OF VARIANCE OF GRADE, SEX, TRAINING AND TEST OCCASION FOR READING VARIABILITY WITH THREE PURPOSES AND EXPOSITORY MATERIALS

Sources of Variation	df	MS	F
Training	1	.10	.05
Sex	1	2.83	1.35
Grade	2 (1)	13.57	6.49*
Occasions	2 (1)	36.23	17.32*
T × S	1	.88	.42
T × G	2	3.04	1.45
T × O	2	.37	.18
S × G	2	2.26	1.08
S × O	2	1.57	.75
G × O	4	.74	.35
T × S × G	2	2.51	1.20
T × S × O	2	3.02	1.44
T × G × O	4	1.62	.78
S × G × O	4	1.49	.71
T × S × G × O	4	1.53	.73
Within (error)	180 (69)	2.09	

* $p < .05$

TABLE 17

ANALYSIS OF VARIANCE OF GRADE, SEX, TRAINING AND TEST OCCASION FOR READING VARIABILITY WITH THREE PURPOSES AND NARRATIVE MATERIALS

Sources of Variation	df	MS	F
Training	1	.86	.30
Sex	1	8.61	3.01
Grade	2	7.87	2.75
Occasions	2	2.55	.89
T × S	1	1.61	.56
T × G	2	1.28	.44
T × O	2	1.70	.60
S × G	2	1.24	.43
S × O	2	28	.10
G × O	4	.20	.42
T × S × G	2	2.34	.82
T × S × O	2	3.62	1.26
T × G × O	4	3.25	1.13
S × G × O	4	1.93	.67
T × S × G × O	4	.69	.24
Within (error)	180	2.86	

The analyses summarized in Tables 16 and 17 also reveal a difference in rate variability across tests – post-training, immediate transfer, and delayed transfer – with expository materials but not with narrative material. Mean rate variability by test and type of material is shown in Table 19. Scheffé *post hoc* tests show a significant difference between the post-training and the immediate and delayed transfer tests with expository material. The suggestion is that not only does the ability to adjust rate to purpose transfer from training passages to more naturalistic materials, but also there is an increase in rate variability. The increase might be expected on the basis that the materials in the

TABLE 18

MEAN RATE VARIABILITY ON ALL TESTS BY GRADE LEVEL AND TYPE OF MATERIAL

Type	Grade 4	Grade 5	Grade 6
Expository	5.98	5.65	5.12
Narrative	5.70	5.93	5.12

transfer test situations were more nearly like classroom materials and less rigidly structured; therefore, the subjects were better able to use their rate variability skill.

It must be noted that the main effects of grade and test were significant only with expository materials. Yet, the trends were generally similar. The structure of the expository material apparently was more conducive to rate variability than that of narrative material. Whether this is generally so, of course, remains to be seen.

There was no difference in the total reading variability of boys and girls. Boys and girls thus learn the same amount during the training situations. No differences were found on any of the interaction factors, which indicates that all factors were consistent in combination.

The type of training materials used did not result in a difference on the total reading variability across occasions regardless of whether the children were reading expository or narrative material. This again emphasizes the similar effect of either of the two types of training materials. It may also indicate that in the future this type of research ought to be concerned with combinations of materials and instructional procedure in order to maximize the effect of training.

TABLE 19

MEAN RATE VARIABILITY FOR EACH TEST BY TYPE OF MATERIAL

Type	Test: Post	Immediate	Delayed
Expository	4.77	5.99	6.01
Narrative	5.37	5.73	5.43

FOOTNOTES

1. The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education and Welfare. This paper first appeared in manuscript form entitled "Transfer Effects of Training Intermediate Grade Pupils to Adjust Reading Speed to Reading Purpose," Cooperative Research Project No. 3137, Committee for Research in Basic Skills, The University of Wisconsin, 1966.

2. An appendix containing the directions and materials used can be obtained from the authors on request.

REFERENCES

Greenhouse, S. W., and Geisser, S. On methods in the analysis of profile data. Psychometrika, 1959, 24, 95-112.

Harris, T. L., Herrick, V. E., Macdonald, J. B., and Rarick, G. L. Experimental development of variability in reading rate in grades 4, 5, and 6. Cooperative Research Project No. 1755, Committee for Research in Basic Skills. Madison, Wisconsin: School of Education, University of Wisconsin, 1965.

Harris, T. L., and Rarick, G. L. The problem of pressure in handwriting. Journal of Experimental Education, 1957, 26, 115-178.

Herrick, V. E., Harris, T. L., and Rarick, G. L. Perception of symbols in skill learning by mentally retarded, gifted, and normal children. Committee for Research in Basic Skills. Madison, Wisconsin: School of Education, University of Wisconsin, 1961.

Herrick, V. E., and Otto, W. Pressure on point and barrel of a writing instrument. Journal of Experimental Education, 1961, 30, 215-230.

Otto, W. The acquisition and retention of paired associates by good, average, and poor readers. Journal of Educational Psychology, 1961, 52, 241-248.

Scheffé, H. A. A method for judging all possible contrasts in the analysis of variance. Biometrika, 1953, 40, 87-104.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Shores, H. J. Reading science materials for two distinct purposes. Elementary English, 1960, 37, 546-552, 565.

Strategies for Concept Attainment in Mathematics

Myron F. Rosskopf
Teachers College, Columbia University
New York, New York

IT WOULD be a mistake for the writer, being trained in mathematics rather than psychology, to do basic research in conceptual learning. On the other hand, it is quite appropriate for him to acquaint himself with the results of experimentation by psychologists and to manipulate these results in different ways to see how they apply in mathematics instruction. In the latter sense, one might say the paper has an operational flavor. That is, it discusses ways classroom teachers might attack instruction in a specific concept of mathematics and means writers of classroom materials might use to insure that provision had been made for learning experiences that would lead to attainment of a specific concept.

A teacher who faces a mathematics class five days each week often finds psychological research of very little help in solving his instructional problems. The research, for obvious reasons, must start with those situations that present knowledge can handle. A research psychologist legitimately pursues his interest in pure psychology; he may have fleeting thoughts about applications of his research, but the excitement of carrying on investigations in his particular field of interest leads him to put these thoughts aside – temporarily, perhaps.

Today, however, there exists sufficient evidence on human learning as well as numbers of well-trained mathematics teachers so that applications can be made in classrooms. It is possible today to have classroom experimentation that makes use of, or manipulates, basic psychological research in human learning. Emerging among mathematical educators and psychologists are plans for studies that will apply in a subject matter field the results obtained from the intensive study of conceptual learning of recent years (that is, since World War II). Every discipline has its own subtleties. For this reason, success of such experimentation will depend upon facility of communication between psychologists and representatives of other disciplines.

DEFINITIONS

CONCEPT

The word "concept" seems to have as many definitions as there are writers on conceptual learning. Recently, a definition appeared that has the virtue of efforts made to relate it to existing definitions. According to this recent statement,

> Anything that is a concept has the following attributes:
> I. Psychological meaningfulness
> II. Intrinsic, functional, or formal properties
> III. Abstractness
> IV. Inclusiveness
> V. Generality
> VI. Structure
> VII. Function (Klausmeier, Davis, Ramsay, Fredrick, and Davies, 1965, p. 3).

A possible criticism is that the definition introduces words which in turn need to be defined. True, but it turns out to be easier to define (or describe) these new words than it was to define "concept."

Two aspects of the definition that make it useful are its open-endedness – by means of item V,

Generality - and its own inclusiveness. Simple concepts as well as very complex concepts are possible of characterization by means of the definition.

The writer has been using the following definition of a concept, probably because it lent itself so well to explanations of mathematical concepts. A concept is "a way of grouping an array of objects or events in terms of those characteristics that distinguish this array from other objects or events in the universe" (Bruner, Goodnow, and Austin, 1956, p. 275). This suffices for a first level concept, but to obtain the complexity that we know, at least intuitively, some concepts have, we need to add to the definition cited; and including "the network of inferences that are or may be set into play by an act of categorization" (Bruner, et al, 1956, p. 244).

We shall not stop to relate these two definitions. Interested readers can go to the references cited and see that the two do indeed contain the same elements. Instead we shall consider an example of a concept in mathematics and show in what way it satisfies the conditions of the first definition stated. Consider, for example, the set of all counting numbers, or as they are often called the set of all natural numbers,

[1, 2, 3, ..., n, n + 1, ...].

One can discriminate certain subsets of this set. There is the set of all even natural numbers,

[2, 4, 6, ..., 2n, ...];

and the set of all odd natural numbers,

[1, 3, 5, ..., 2n + 1, ...].

For our example of a mathematical concept, however, we shall consider that of the set of all prime (natural) numbers,

[2, 3, 5, 7, 11, 13, ...].

When a person can distinguish between prime numbers and numbers that are not prime (that is, that are <u>composite</u>), then he has a concept of prime numbers. He can demonstrate his possession of the concept by filling in the next few prime numbers, or by devising a test for deciding whether a number presented to him is prime, or not prime. He has grouped the array of objects, the set of natural numbers, so that he can tell the difference between a prime number and a composite number. He has used the characteristics that distinguish between prime numbers and other members of the set of natural numbers. Really, the person has formed two <u>categories</u> of natural numbers; those that are prime form one category; and those that are composite form the other category.

At this stage of his concept of a prime number, a person recognizes the <u>formal properties</u>. He certainly is working at an <u>abstract level</u>, since he is working with numbers, and at a level of <u>generality</u>, since many subsidiary concepts such as factorization, multiplication, and so on are involved. From the point of view of <u>structure</u>, the person probably has used a conjunction of related concepts to arrive at a concept of prime numbers.

But we recognize more in a concept than just the forming of categories according to some discriminating characteristic or characteristics. Part of the concept of prime numbers is realization that their number is infinite, that there is just one even prime, that no one has devised a formula for prime numbers, and that there are many such open questions related to this interesting subset of the set of natural numbers. In other words, we are willing to accept as part of the concept these inferences drawn from the categorization.

It seems that such associations with the set of prime numbers would qualify it for the attribute of <u>psychological meaningfulness</u>. There remain the attributes of <u>inclusiveness</u> and <u>function</u>. Now, inclusiveness is so much a part of specifying a set in mathematics that as soon as one is able (that is, correctly in the sense of mathematical criteria) to talk about the set of prime numbers, the attribute of inclusiveness follows. Prime numbers as mediators occur in prime factorization, calculation of a least common multiple, and the prime factorization theorem of arithmetic. Hence, the concept has the function attribute also.

All right. It is possible to relate what mathematicians call a concept to what psychologists call a concept. But a definition, if it really is a definition must partition the universe of discourse into those things that are concepts and those things that are not concepts. A "definition" that is all inclusive might just as well not be stated at all. In short, it must be possible to point to something in mathematics that is not a concept. But that is easy; any addition fact, such as 5 + 4 = 9, is not a concept.

CATEGORIZING OR CONCEPTUALIZING

The word "category" was used in the preceding discussion. What is involved in categorizing? How is it related to conceptualizing? More important, how is categorization related to the use of conceptualization in applications, or problem solving, in mathematics? First, let us analyze another example of a mathematical concept and, second, derive from the discussion of the example what seem to be characteristics of categorizing.

Consider the concept of a rational number. Its formation starts very early in the home experiences of a preschool child. One-half and one-fourth are known for names of parts of things; one-half of an apple or an orange or a glass of juice or a candy bar. During early schooling, the concept is expanded to include written symbols for one-half and one-fourth: 1/2 and 1/4; 1/3 also is added to the list. However, the concept is still on the level of these symbols being names for parts of things. There seems little evidence that a child thinks of 1/2, 1/4, 1/3 as exemplars of a new sort of number. The teacher may call these fractions names for numbers, of course, but still the concept, the way of categorizing these numbers, of telling how they differ from other numbers, is in terms of a way of expressing them; to wit, a numerator, a horizontal bar, and a denominator.

Not until quite late does the instruction reach the stage of giving a new name, rational numbers, to the set of numbers that can be expressed as the quotient of two integers, the denominator not zero. Now the concept has formal properties; it has been formalized, one might say, and it will be even more formal (that is, abstract) when a student recognizes

rational numbers as ordered pairs of integers in which the second member is not zero.

At each level of formality, there are cues that a student looks for: 1) one of two congruent parts of something, such as an apple; 2) a numerator, a horizontal bar, and a denominator; 3) a quotient of two integers; and 4) an ordered pair of integers. These cues help a student put a number he encounters into its proper category, to align it with the concept he has formed of the set to which that number belongs. Notice, too, that the invention of ways of grouping predictive characteristics is involved; for example, the numerator, horizontal bar, and denominator. The test of the invention is its usefulness. Does it effectively categorize numbers?

Categorizing has for its basis defining characteristics. By grouping together things that satisfy the defining characteristics of a class, an individual a) reduces the complexity of his environment; b) identifies the objects in a particular universe of discourse; c) reduces the necessity for constant learning; d) gives direction to activity; e) orders and relates classes (Bruner, et al., 1956, pp. 12-13).

What are the implications for problem solving in mathematics? If a student does not classify problems he meets, then he must treat each problem as an individual. The cognitive strain in such a procedure is enormous. The student must, perforce, for purposes of economy of effort look for distinguishing characteristics or discriminable attributes of problems in order to reduce the strain. That is, he must use cues, put the problem in a pre-decided-upon category, and bring into play concepts that in the past have been effective in solving problems of this particular category. It would seem that a categorization of a problem is the first step toward its solution. Only after a problem has been put in a certain category can one proceed with plans for solving it.

Of course the first plan chosen may fail. If it does, a person has to devise variations on the first plan. Perhaps he failed to discriminate carefully enough when he categorized the problem. In other words, taking a first step toward its solution, leads him to recategorize the problem. The important thing is that he must realize that the initially chosen method of solution fails and then search for a new method. He must shift his plans, and this requires strategy.

STRATEGIES IN CONCEPTUALIZATION

STRATEGY

A strategy is a carefully planned calculation and coordination of the specific ends and means necessary to achieve a goal. That is, as in the following:

> A strategy refers to a pattern of decisions in the acquisition retention, and utilization of information that serves to meet certain objectives, i. e., to insure certain forms of outcome and to insure against certain others. Among the objectives of a strategy are the following:
>
> a. To insure that the concept will be attained after the minimum number of encounters with relevant instances.
> b. To insure that a concept will be attained with certainty, regardless of the number of instances one must test en route to attainment.
> c. To minimize the amount of strain on inference and memory capacity while at the same time insuring that a concept will be attained.
> d. To minimize the number of wrong categorizations prior to attaining a concept. (Bruner et al., 1956, p. 54.)

The preceding definition of a strategy and a description of the aims of a strategy is very similar to the mathematical definition of a strategy that is associated with the theory of games. However, the theory of games strategy consists of a set of decisions made in advance. Once the decisions are made, no change can be made while "play" is going on. In contrast, a problem-solving strategy, for example, may be changed in the midst of a subject's efforts; such changes are dictated, or should be dictated, as the subject's knowledge of the problem grows through his efforts to find a method of solution.

TYPES OF STRATEGIES

"Virtually all the effective strategies for attaining concepts depend upon the use of some sort of initial focus," (Bruner, et al., 1956, p. 63). How does the subject perceive the situation? What aspects of it holds his attention? How does he define the task to himself? If the task is a problem situation, what data seem to be most relevant? How does he fit this problem into his concept pattern of all types of problems with which he has had experience? After making an initial focus with respect to the problem, a subject still faces the necessity for decisions concerning a method of attack. What sort of strategy should he decide upon? According to Bruner and his associates there are four ideal strategies: a) simultaneous scanning; b) successive scanning; c) conservative focusing; and d) focus gambling (Bruner, et al., 1956, pp. 83-90).

a) Simultaneous scanning. The simultaneous scanner forms many possible hypotheses to follow to a solution and keeps all of these hypotheses in mind. The strain on memory is great. Initially, he chooses one of these hypotheses and tests it to see if it is promising and at the same time to see which of the other hypotheses it eliminates. A subject using this strategy has to keep all of the hypotheses in mind; he has to deduce which are eliminated, and as he proceeds to test a remaining hypothesis, he must remember which hypotheses were eliminated by his earlier testing.

The critical point in simultaneous scanning depends upon a subject's ability to deduce all the information with respect to which hypotheses should be discarded from the testing of an initial hypothesis. If he can do this, then he can proceed with maximum information to test a remaining hypothesis. However, memory strain and the factor of time usually prevent him from obtaining maximum information. Suppose the subject makes a wrong deduction? Does he have control over this risk? He does not; he cannot himself either increase or decrease the risk.

b) Successive scanning. Essentially, this strategy consists of testing one hypothesis at a time. A hypothesis is tested until a contradiction is reached, or until it is clear to the subject that the desired information is not forthcoming. Then the hypothesis is discarded, and a second is chosen for testing. Notice that a hypothesis is tested independent of the others.

The gain in use of successive scanning is twofold: 1) it lightens memory load and 2) frees the subject from the necessity for making many deductions concerning the other hypotheses. Two disadvantages are that some of the hypotheses may overlap, leading to repetition of work already done, and that there is no control over risk. A subject cannot take greater gambles or lesser gambles. This is a rather leisurely strategy and discounts heavily any time factor.

c) Conservative focusing. First of all, a subject who is a conservative focuser studies the situation to see if there is one key idea. If there is, it is focused upon, and all hypotheses are tested against it. Suppose he finds a hypothesis that looks promising. That is, a hypothesis that impinges on the key idea. Now, he chooses those hypotheses that contain the same part that the promising hypothesis does. Then, one by one, he adds parts of these hypotheses and tests the results. The "conservative" enters this strategy at this point. Only one feature of a hypothesis is added at a time and tested.

Much information is obtained by conservative focusing and redundancies are avoided. The risk of securing no information is under control; in fact, one always gets some useful information. The only real disadvantage is that information which is not arranged in orderly fashion might involve a long search for the needed data. Thus, part of the success of a subject using conservative focusing depends upon how well organized his knowledge is.

d) Focus gambling. The initial steps taken by a subject who uses focus gambling as a strategy are the same as those of one who uses conservative focusing. But, he is not satisfied with adding one part of a hypothesis at a time and testing; he adds two or more parts at a time and gambles. The strategy is not quite an all-or-nothing one, but it comes very close in case the omnibus hypothesis fails. The only thing to do then is for the subject to shift strategies and fall back on successive scanning.

However, the gamble may pay off. The omnibus hypothesis may lead to attainment of the objective. There is much gain then, of course, for the number of trials has been reduced to a minimum and much time has been saved.

THE STRATEGIES IN USE

IMPORTANCE OF INITIAL FOCUS

Several years ago, the writer did a pilot study in which about ninety secondary-school teachers were involved. Two tasks, or problems, were presented to the subjects, and they were asked to find a solution. The objective at the time was to secure information on these teachers' concept of a mathematical solution, but other interpretations of the evidence turned out to be more interesting. This is true, particularly in view of the foregoing discussion.

Situation 1. One problem was the following:

From a two-digit number, N, subtract the number whose digits are those of N interchanged. If the difference is a positive cube, how many permissible values of N are there?

However innocent this problem appears, it is not a run-of-the-mill problem. The method of approach, that is, the initial focus, is critical. Practically all of the teachers classified the problem as a "digit problem." All right; there are standard methods for solving a problem so categorized. These were applied; let t and u represent the tens' and units' digits, respectively; then

$$N = 10t + u$$

a positive perfect cube $= (10t + u) - (10u + t)$.

Quickly perceived is that the tried-and-true methods fail. There is no place to go from this point, for at least two more equations are needed for them to be successful. A small number of papers were discarded because of errors in finding a simpler form for $(10t + u) - (10u + t)$. It is doubted that this was failure of any strategy. Rather, such errors seemed to be the result of haste or carelessness. The remaining papers fell nicely into three groups.

There were those whose initial focus was all right. These papers showed that the persons got as far as

$N = 10t + u$
$x^3 = 9(t - u)$
$x > 0$.

But, unfortunately, once they had made a focus, they were not flexible enough to change it when it led nowhere. The papers of these subjects were covered with equations and tag ends of scribblings. They formed only one hypothesis – secure enough equations to solve for all of the unknowns. One might categorize their strategy as successive scanning, where there was just one hypothesis to scan. This was done over and over by members of this group. There was no change of focus with respect to the problem. Failure did not discourage them from trying the same hypothesis again and again.

The second group's papers showed fewer scribblings. There was hardly any writing on these papers after the initial symbolizing of the problem situation. Inferred from what evidence was available was that these subjects realized that the initial categorization of the problem as a digit problem was not quite correct. They seemed to be flexible enough in their focusing on the problem to change their focus. It seemed, however, that they were at a loss to formulate another hypothesis or another strategy. A possible interpretation is that these people had a limited concept of a mathematical solution. Geometric proofs and the results obtained by solving equations qualified all right, but the result found by trial and error or by a verbal argument did not.

Members of the third group not only realized that a shift in initial focus was necessary but they also had a better conception of what constitutes a mathematical solution. Some adopted a strategy of

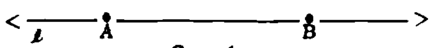
successive scanning, trying different pairs of numbers in succession. Others clearly used conservative focusing, for they not only tried pairs of numbers but also took into account the positive perfect cube datum. Those with the most elegant solutions used a highly sophisticated conservative focusing strategy. A few used focus gambling, but more due to the pressure of time rather than to preference.

Situation 2. The second task presented to these teachers was a straight-forward geometry problem:

Find, on a straight line, a point such that the sum of its distances from two given points is the least possible.

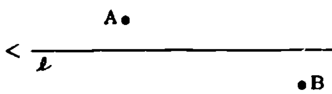
The first step in any geometry problem is to draw a figure and introduce some notation. If more subjects had followed this advice, there would have been more correct solutions to this problem. As it was, only one-sixth of the subjects were able to arrive at a solution. Just as was the case with the algebra problem, here, too, the difficulty was inflexibility with respect to the initial focus.

Let us take the key idea of the problem, the sum of the distance from two given points is the least possible, and use it as an initial focus. Now, as one proceeds to make a figure and introduce notation, he must give



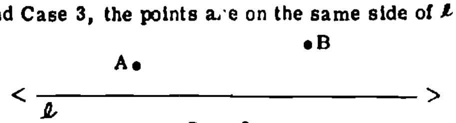
Case 1

up this initial focus for a moment and shift to focusing on the two given points. The problem says nothing about placing them; a decision is required. Three possibilities, or hypotheses, appear: Case 1, both points are on line ℓ ; Case 2, the points are on opposite sides of ℓ ;



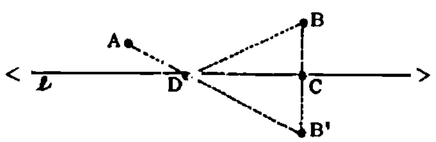
Case 2

and Case 3, the points are on the same side of ℓ .



Case 3

A shift of focus back to the least possible distance-focus immediately takes care of Case 1 and Case 2, for the shortest distance between two points is the length of the segment joining them. Now, keeping these results in mind, turn to Case 3 and ask "How can I use this idea (the length of the segment)?" A few moments' thought suggests reflecting one of the points in the line ℓ , yielding point B'. Segment $\overline{AB'}$



intersects ℓ in D, and D is the desired point. For triangle BDC is congruent to triangle B'DC, and consequently BD = B'D. Therefore, BD + DA is the least possible distance.

These two problems have emphasized the importance of initial focus. More than that, discussion of attempts at solutions made clear how necessary it is for a problem solver to be flexible enough to shift his focus as the process of solving the problem develops. In each problem, concepts related to that involved in the problems themselves were very important. It seems that, at least in mathematics, knowledge of these subsidiary concepts is crucial to success. The strategies used seemed first to be that of simultaneous scanning with a transfer to conservative focusing after a promising "lead" had been discovered.

SIMULTANEOUS OR SUCCESSIVE SCANNING?

The next problem situations are relatively independent of initial focus, depending much more upon a choice of strategy. The mathematics curriculum puts considerable emphasis today on a search for patterns. The following situations are of that sort.

Situation 3. Divide your paper into three columns. In the first column there is a string of letters, and in the second column there is another string of letters. Now, we shall use these two strings of letters to make a new set

K, L, M, T, W, Z	L, M, T, X, Y	

of letters in the third column. How were the letters in the third column

K, L, M, T, W, Z	L, M, T, X, Y	L, M, T

obtained? Now, try to write the set of letters that goes into the third

A, C, E, P, Q	A, C, L, P, Q, R, S	

column for these two sets. What goes into the third column for the following two sets?

T, U, V, W	T, U, V, W	

Here is another example. What would you write in the third column?

S, A, M, E	N, O, T	

Is it true that, given sets in the first and second columns, there is always exactly one way to do the third column?

What are the results if we interchange the sets in the two columns? Fill in the third column and compare the results with what you got before.

L, M, T, X, Y	K, L, M, T, W, Z	
A, C, L, P, Q, R, S	A, C, E, P, Q	
T, U, V, W	T, U, V, W	
N, O, T	S, A, M, E	

What did you find?

Situation 4. Now let us form a third column set in a different way. As before, we start with sets in the first and second columns. How was the

A, C, E, G, H, J	B, C, D, E, F, G, T	A, H, J

third column set obtained? Now, try to write the third column set, starting

F, A, M, I, L, Y	E, A, R, L, Y	

with these two sets. Let us try two more special situations. Is it always

A, B, C,	D, E, F, G	
A, B, C, D	A, B, C, D	

possible to obtain exactly one set, given sets in the first and second columns?

Let us see what the results are if we interchange the sets in the two columns. Compare the answers with those you had before. What can you say?

B, C, D, E, F, G, T	A, C, E, G, H, J	
E, A, R, L, Y	F, A, M, I, L, Y	
D, E, F, G	A, B, C	
A, B, C, D	A, B, C, D	

We have worked with strings of letters in two different ways. That is, we used two different patterns to obtain the set in the third column. Compare these two patterns. Do they have the same properties? How are they different?

DISCUSSION OF SITUATIONS 3 AND 4

The last two problem situations are quite different from the first two. Except that they were not quite straight-forward, the first two problem situations are representative of the mathematics program before World War II. The last two are representative of types of problems that arise in the new mathematics program. The algebra and geometry problems have for their objective getting an answer. After the answer has been obtained, there is little one can do with it, except perhaps to refine the method of solution. It is not so much that a concept is in the process of being attained as it is that pre-established concepts serve as mediators in solving each problem. Consequently such problems would be reasonable to present only after there was some evidence that subjects had been at least exposed to the necessary mediating concepts.

On the other hand, the third and fourth problem situations are representative of situations intended for seven- to nine-year-old subjects. The intent is that through such situations a foundation is laid for attaining rather deep mathematical concepts – those of closure and commutativity in this case. A battery of situations of this sort can be devised that has a bearing on ordinary arithmetic. Of course, letters would not have to be used. One could use the boys and girls in a class, or colored counters, or books, or any other sort of material available. An effort would be made to use materials appropriate for the age level of the class.

It is not clear to the writer that mediating concepts are brought into play. Rather, it seems that solving these problem situations requires use of a strategy, one that is effective in attaining concepts. Pattern-problems such as these seem to yield to a successive scanning or a simultaneous scanning strategy. Or, one uses a combination of the two strategies; simultaneous scanning for gross eliminations of parts of the strings of letters; and successive scanning for determination of fine discriminations.

TWO ITEMS OF RESEARCH

Important to the mathematics program is knowledge concerning what types of data in problem situations present difficulties. Such knowledge influences preparation of classroom materials in two ways. Learning experiences can be provided that are intended to facilitate attainment of a concept, and information gleaned from research studies can be used to schedule concepts according to age levels.

An attempt by one person to report on all of the research studies concerned with problem solving would be foolhardy. A bibliography is very long. Two doctoral studies have been singled out for reporting. One reason is to underscore the fact that much relevant research gets lost or overlooked through non-publication in journals.

POST STUDY

In 1958, Post did an experimental study of six factors involved in understanding problems. His subjects were fifth- and sixth-grade pupils in schools in the Metropolitan New York section of the country. The definition of a problem in the study was "a quantitative situation described in words ... but in which the arithmetical operations that lead to the answer to the question are not explicitly indicated" (Post, 1958, p. 15). The six factors were: 1) size of numbers; 2) superfluous numerical data; 3) familiarity of setting; 4) number of steps; 5) type of operation; and 6) symbolic terms.

Factor 4, number of steps, requires some words of explanation. Post defined a two-step problem and a three-step problem as follows:

> A step in a problem solution ... was a simple performance of one of the four operations of addition, subtraction, multiplication, and division. Therefore, a problem is (a two-step problem) if and only if the solution of the problem requires a single performance of two different operations or two performances of the same operation (Post, 1958, p. 30).

A three-step problem is defined in a similar way.

Factor 6, symbolic terms, refers to the question of whether using a name for a number such as "twenty-six" in place of "26" interferes with a subject's ability to solve a problem. The names of the other factors are self-explanatory.

Two forms of a test were prepared by Post; these forms were judged equivalent by a jury of experts and then tested further in a pilot study. Familiarity of setting and non-familiarity of setting was obtained from an inventory of such settings suggested by elementary-school teachers. This information was

supplemented with evidence obtained from the pilot study pupils.

The study design was such that an analysis of variance technique could be used. The error term "was composed of the unconfounded four- and five-factor interactions as well as the six-factor interaction" (Post, 1958, p. 85). The two forms of the test were treated as two replications of the same experiment, for examination of the test results from the two forms led to doubt that the data from them could be combined. The schools represented two different socio-economic neighborhoods; the subjects were chosen from non-homogeneous classes; and small sample theory analysis was used. Correctness of solution was determined from examination of the work written by a pupil rather than from the correctness of the answer. For example, a pupil's solution might be categorized as correct even if his answer was incorrect. Another way of saying this is that errors in arithmetic computation were disregarded.

A hypothesis was rejected if the results from both forms of the test indicated rejection; it was accepted, similarly, if both forms indicated acceptance. No conclusion could be stated if one form indicated rejection and the other form indicated acceptance.

The hypotheses on the factors of size of numbers and number of steps fell in the no conclusion category. So far as the latter (number of steps) is concerned, it might have been that there was a strong interaction between number of steps and type of operation, for the hypothesis concerning the latter was rejected at the one tenth of one percent level of significance. That is, problems involving multiplication and division were found to be more difficult for the sample subjects to understand than problems involving addition and subtraction.

Hypotheses rejected were those concerning superfluous numerical data, familiarity of setting, and the already mentioned type of operation. The hypothesis concerning symbolic terms was accepted. In short, expressing numerical data in words or in digits had no effect on pupil's understanding of such problems.

BECHTOLD STUDY

The 1965 study of Bechtold consisted of a comparison of an experimental group with a control group in which the former received instruction in work with problems containing superfluous data and the latter did not. Two units from the ninth-year program, inequalities and applications of equations to falling body problems, were chosen for instruction and testing. The subjects were 14- to 15-year-old students in New York City and the surrounding Metropolitan region schools.

The hypothesis that instruction in handling problems containing superfluous data (in both inequality and falling body problems) had no effect on developing problem-solving ability was rejected at the one percent level of significance. The results with respect to variables of sex, intelligence, and age indicated no significant differences with respect to sex and intelligence. The data indicated, however, that younger students seemed to benefit more from instruction than older students.

RELATION OF THE POST AND BECHTOLD STUDIES TO CONCEPT ATTAINMENT

It is clear that if there is such a thing as a concept of problem solving, then it is a very complex concept. Moreover, success with problem solving is dependent upon mediating concepts that in themselves are complex. The studies of Post and Bechtold, for example, underscore the importance of instruction, of exposure of pupils to many different sorts of problems or learning experiences in order that they have opportunities to attain necessary subsidiary concepts.

The fact that familiarity of setting and superfluous data were significant factors in Post's study seems to indicate that pupils may have had no previous encounters with such problems. Hence, they would have had no opportunity to form a concept of such problems, that is, to categorize them and develop strategies to use in their solution. Since multiplication and division in problems turned out to be significantly more difficult to understand than addition and subtraction, one infers that pupils do not conceptualize the former as well as the latter. It may be that the methods of teaching used with multiplication and division failed to develop firm concepts. As a result, pupils were unsure in their categorizations when these operations were present in problems.

Bechtold established that instruction in how to handle problems containing superfluous data led to greater success with such problems than that of a control group - at least for ninth-grade algebra students. A teacher is prone to extrapolate from this evidence and conclude that such instruction, appropriately designed, is good for almost all age levels. Although the study does not represent conclusive evidence, still it is encouraging in that statistically significant differences resulted. So often an experiment that compares two procedures in instruction ends with one group doing about as well as the other on tests; there are no statistically significant differences in the two sets of results. The unexpected significant differences in Bechtold's study leads one to go beyond the statistics and say that experience with the analysis of problems with superfluous data significantly affects outcomes.

From the definition of a concept given at the beginning of this paper, one sees that a person does not form a concept just through maturation. There must be experiences in his environment that make his attempts to understand the environment too complex unless he forms classes - unless he conceptualizes. This is particularly true of mathematics, one of man's sophisticated inventions. Perhaps a child develops "naturally" some concept of counting, even inventing his own names in some instances. But as there are encountered larger and larger groups, the effort to remember number names becomes too much, and, if historical evidence is valid, there is a turning toward forming classes, groupings, and a numeral system is born. From this point on, present day culture dictates

that intensive exposure to a variety of experiences must leap-frog the historical pattern of development in mathematics. These learning situations cannot be left to chance; they must be provided; and they must be planned carefully. Experimental evidence, both in the field of psychology – learning theory – and mathematical education, furnishes guidelines for such planning. In turn, experimental evidence affirms or denies that the planning is effective. The Post and Bechtold studies are examples of this two-pronged approach to a mathematics curriculum.

SUMMARY

So it is that human beings tend to deal with classes of things instead of individuals in order to make some sense out of their environment. By forming such classes cognitive strain is reduced as well as the burden on memory. To form these classes, or categories, or sets, a person looks for cues, or if you like, for characteristics that serve to distinguish things eligible for membership in the set from those that are not eligible. The point is that these categorizations are inventions, and this is particularly true in a complex body of knowledge like mathematics.

To solve a problem in mathematics, subsidiary concepts serve as mediators. Although initial focus and type of strategy have important roles, an individual will have little success in problem solving unless he has a firm grasp of the mediating concepts. Mathematical education speaks in terms of a student mastering a concept so that it is a part of himself; a psychologist would speak in terms of a student internalizing a concept. No matter what language one uses, it is certain that some of the failures in mathematics instruction are due to an instructor assuming his students have understood a mathematical concept at a high level of operational thinking when in reality they have a much lower level of mastery. They cannot apply the concept in an unfamiliar setting. As experimental evidence accumulates from investigations of how people learn and from studies in a curriculum context, mathematical educators are on the one hand becoming more cautious and on the other more daring. Applying the evidence, they have changed the content of the curriculum and reduced the amount of pure drudgery that used to be the lot of every mathematics student. Assessing recent evidence, they have some doubts whether all students master the early introduction of concepts.

Already said is that people categorize the things in their environment. A categorization might result in an initial focus on a problem. Then certain hypotheses are formed that are related to this initial focus. Decisions on these hypotheses serve to validate the hypotheses.

> The sequence of decisions made by the person en route to attaining the concept [which may be a problem solution or a pattern] ... may be regarded as a strategy embodying certain <u>objectives</u>. These objectives may be various in kind but in general one may distinguish three kinds of objectives: <u>a</u>. to maximize the information gained from each decision and test of an instance; <u>b</u>. to keep the cognitive strain involved in the task within manageable or appropriate limits and certainly within the limits imposed by one's cognitive capacity; and <u>c</u>. to regulate the risk of failing to attain the concept within a specifiable time or energy limit... (Bruner, et al., 1956, p. 234).

In the examples presented of problem solving or of attaining a concept in mathematics, the idea of initial focus and flexibility with respect to one's focus is of great importance. After isolating a key idea in a problem, a student has a choice of strategies – simultaneous scanning, successive scanning, conservative focusing, or focus gambling. A skilled problem-solver probably uses a combination of simultaneous scanning and conservative focusing. A beginner is likely to use successive scanning. If time harasses a problem-solver, there is nothing to do but turn to focus gambling as a strategy. The situation and the individual (his own experiences, battery of concepts, and psychological make-up) determine which strategy will be used.

The investigations of Dienes (1963) at Harvard as well as that of Dienes and Jeeves (1965) in Australia lean heavily on the work of Bruner and Piaget. However, their work goes beyond that of their predecessors, making use of new research results and suggesting refinements of the theory of learning upon which Bruner based his work. An idea that Dienes and Jeeves, separately and together, experimented with is the development through "mathematical games" of the concept of isomorphism of structures. Involved here is the recognition of patterns and noticing the similarity between two sets of data and the structure that can be applied to both.

Suppes (1965) has done a great deal of work with the major emphasis on learning models for relatively non-complex mathematical concepts. He characterizes his work by saying that it is based upon a stimulus-sampling theory. An interesting assertion is that the hypothesis-strategy language of Bruner and his associates can be related to stimulus-sampling terms: "... a strategy ... corresponds precisely to a state of conditioning and a hypothesis to the conditioned stimulus sampled on a given trial ..." (Suppes, 1965, p. 65).

Some psychologists doubt the value of a mathematical psychology approach to models for human learning, but there is no disagreement that much work needs to be done with complex concepts in order to learn more about individually different approaches to attaining mathematical concepts. Meanwhile, the mathematical educator makes use of available knowledge, applies it as best he can to the mathematics program, and gathers experimental evidence to see the effect. It is impossible to wait until definitive psychological research on learning has been done, for the curriculum exists and must be modified to fit twentieth century activities. One implication is clear: new psychological evidence should be brought to bear on the curriculum as quickly as possible. It is in the latter direction that mathematical education is turning.

NOTE: Situations 3 and 4 are modeled after similar materials that appeared in the <u>Nuffield Foundation Mathematics Teaching Project Bulletin</u> No. 1, November 1964, pp. 4-7.

REFERENCES

Bechtold, C. The use of extraneous material in developing problem-solving ability of pupils in Algebra I. Unpublished doctoral dissertation, Teachers College, Columbia University, 1965.

Bruner, J. S., Goodnow, Jacqueline J., and Austin, G. A. A study of thinking. New York: Wiley, 1956.

Dienes, Z. P. An experimental study of mathematics-learning. London: Hutchinson, 1963.

Dienes, Z. P., and Jeeves, M. A. Thinking in structures. Psychological Monographs on Cognitive Processes. Vol. 1. London: Hutchinson Educational, 1965.

Klausmeier, H. J., Davis, J. K., Ramsay, J. G., Fredrick, W. C., and Davies, Mary H. Concept learning and problem solving: A bibliography, 1950-1964. Technical Report No. 1, Research and Development Center for Learning and Re-Education, C-03, OE5-10-154. Madison: University of Wisconsin, 1965.

Post, R. A study of certain factors affecting the understanding of verbal problems in arithmetic. Unpublished doctoral dissertation. Teachers College, Columbia University, 1958.

Suppes, P. On the behavioral foundations of mathematical concepts. In L. N. Morrisett and J. Vinsonhaler (Editors), Mathematical learning. Monographs of the Society for Research in Child Development, 1965, 30, (1, Whole No. 99), 60-96.

PART III

Instructional Materials and Media

THE MEDIUM of instruction can be considered as the interface between the learner and the subject matter being learned. It necessarily affects the rate of learning and imposes limitations on the scope and sequence of the instructional subject matter. Instruction by television has been criticized as impersonal and encouraging passivity. Yet, according to Wilbur Schramm in "Instructional Television Around the World" instructional television can, when properly used, provide for active learning experiences. In a series of case studies, Schramm reports on the use of television in widely varying sections of the world—from the affluent suburb to the developing nation. Properties and attributes of the media are discussed with reference to the school system, the teacher, and the student. Television is examined as a potent mechanism of change. Schramm concludes that this medium works best when it is integrated into an instructional system which includes specialized materials, activities, and teaching techniques. But the key to the system is a local supervisor through whom students can interact with one another.

Research dealing with the process of instruction is examined by Arthur Lumsdaine who points out that problems associated with instruction are in fact inseparable from problems dealing with learning. The close relationship between research on instructional media and the more general field of basic psychology of learning must be understood by persons purporting to do research on instructional media. Among the general areas identified for research on the effects of instruction is the "far underdeveloped research field of constructive and imaginative attempts to improve on the basis for determining what ought to be taught." Lumsdaine argues that determining the effects of present instructional media should precede the "more elegant question of what specific variables account for these effects." He takes the position that effective instructional media are a prerequisite to meaningful basic research on these media. This implies programs of development before research and suggests a basis for priorities in research and development. "Only when something effective has been developed, do we then have an adequate tool with which to conduct useful research."

The question of cost efficiency which is usually not considered by educators is explored. Such efficiency frequently implies a measure of the level of proficiency achieved for a given effort with, as Lumsdaine suggests an approach in which the time required to achieve a given minimum level of proficiency is used as the dependent variable. Such a proposal has significant implications for the design of instructional objectives and instructional research as applied to selected subject areas.

Lumsdaine concludes with a discussion of the standards or lack of standards used in the design and reporting of research studies dealing with instructional media.

Matching the instructional strategy to the capabilities of a learner requires a tutorial model of instruction. Lawrence Stolurow proposes an instructional system in which methods and materials are individualized for the learner on the basis of the continuous matching of methods and materials with the aptitude, personality, and knowledge of each learner through the use of feedback contingencies both prior to and during instruction. Such individualization can be achieved only in Computer Assisted Instruction. Stolurow summarizes a general idiographic contingency model of a cybernetic, computer based instructional system called SOCRATES. "It (this model) is based on the assumption that individualized instruction should take a prescriptive-corrective approach to the use of psychological information in making decisions about teaching in a tutorial context." The logic of instruction is determined according to certain rules whose application is contingent on variables monitored throughout the instructional sequence. Results based on the idiographic model are presented with reference to the kinds of decisions made in individualizing instruction via the computer.

Instructional Television Around the World

Wilbur Schramm
Stanford University
Stanford, California

WE HAVE now in hand the largest body of comparable and detailed data yet assembled on instructional television throughout the world: twenty-three case studies of the use of new educational media in eighteen countries, most of them including television.[1] I will review three of these cases in some slight detail, cite a few others, and then try to sum up what I think the studies say about the requirements for using instructional television effectively.

One point I shall not belabor is that instructional television, used well, clearly does work. It works over a wide spectrum of tasks and a variety of situations. It works in developing countries and countries that are highly industrialized. It works sometimes very well and sometimes ordinarily. To illustrate what these cases say about how it works and under what conditions it appears to work well, I will enumerate three instances.

CASE 1: SAMOA

The story of instructional television in Samoa really begins with the appointment of a tough, determined, little governor, H. Rex Lee, in 1961, but the whole story is 60 years older than that, and deals mostly with American neglect of its wards on the Samoa islands. The school system was only one symptom of that neglect, but it was a vivid one.

The United States had committed itself to offering Samoans an education equal in every way to that on the mainland. Yet the Samoan schools were in one-room, open fale huts, where two classes competed with each other, fortissimo. The instruction itself was the most traditional rote learning. I have been in these village schools and have heard the kind of choral chanting that one becomes accustomed to in primitive schools throughout the world. The students repeat back – chant back – what the teacher gives them; indeed, this is the only kind of teaching possible when the teacher is untrained, little educated, and not at all confident of his grasp of the subject matter. He doesn't dare encourage questions or let the students wander away on their own intellectual excursions. The teacher says, "Class, what am I doing?"

Like one, the class chants back, "You - are - standing - on - the - mat."

"Class, what do I have in my hand?" the teacher questions.

"You - have - a - piece - of - chalk," the class replies.

Once I rose, smiled as disarmingly as possible at the teacher, and said "Class what am _I_ doing?"

"You - are - standing - on - the - mat," they responded promptly, although I was far away from the mat.

I picked up a book. This puzzled them; they could see it wasn't chalk. "What do I have in my hand?"

One boy ventured: "You - have - a - piece - of - chalk."

They were quite unable to generalize much upon the responses they had learned in their educational catechism.

The results of this kind of thing were predictable. The schools were supposed to be the equals of main-

land schools, but children graduated from the twelfth grade testing on the average from fifth to ninth grade in different subjects. The teaching was supposed to be in English, but little of it was, and many of the teachers could not speak the language understandably. The content was derivative from mainland texts, and so the children learned their reading vocabularies from accounts of commuter trains, milk men, and snowball fights. I saw one elementary-school class being drilled to remember that "The 'Mona Lisa' is in Milan," although the significance of that knowledge to a Samoan child was not immediately obvious. And over everything hung the pall of traditional rote instruction.

Now, how do you transform a traditional school system like that, in a few years, into a modern problem-solving, inquiring school system? How do you live up to the promise and plan to teach in English, with so few English models? That is the problem Governor Lee faced. I am begging the question of whether the level should have been raised in a few years, rather than a century; whether it was important to have English well taught and well used; whether an excellent 12-grade school system was what Samoa most needed. But given that policy goal, how would you proceed?

What were Governor Lee's alternatives? He could replace all three hundred Samoan teachers by qualified teachers from the mainland, if he could recruit them. But this would have been a politically, psychologically, and ethically impossible solution. He could bring in perhaps one hundred teachers from the mainland. This would displace less than one-third of the native teachers and stir up less trouble, but would run head on against one of the oldest Samoan norms - that all go forward together. He could institute a greatly increased teacher program. But this would require sending prospective Samoan teachers to the mainland for some years - lest they merely perpetuate the quality of the old school system - and its effect would not be felt for perhaps 15 years. Or he could use television to share a small amount of expert teaching over many classrooms. And this is what he decided to do.

But television itself was not enough. The people who studied the problem saw very soon that what they had to create was really a new teaching-learning system. They had first to consolidate the schools and provide adequate learning conditions. Then they had to create a sort of teaching team, in which an expert teacher would give the core of the curriculum by television and a native classroom teacher would be taught to build a context of learning activities around the broadcast; and still another teacher would prepare study materials and exercises to fit into the plan. Make a mental note of this, because the same conclusion is being reached all over the world - that whereas pupils learn a great deal from television alone, they learn a great deal more when it is built into such a teaching-learning system as I have described. Nowhere in the world is television being used entirely alone and effectively for any major instructional task. Always it is integrated into some kind of system in which the studio teacher, the classroom teacher, and sometimes others cooperate in stimulating the maximum amount of learning activity in the students.

Governor Lee persuaded Congress that our neglect of Samoan education should be repaired. A remarkable installation was built at Pago Pago, one of the finest instructional television installations in the world, with six open-circuit channels, ten videotape recorders, transmitters on a mountain top, and television teachers, engineers, and producers recruited from the States. Classes began to move into the new school buildings in 1964, and television began to carry the burden of teaching beginning in the autumn of that year.

How is it working? What evidence is there that television can raise a primitive to a modern system - Beeby's (1966) Class I and II schools to his Class III and IV schools - in a short time?

The evidence isn't all in, yet, but the signs are encouraging. The Samoan elementary schools are working very well, indeed. The children are speaking much more and much better English. The television classes are live and active; the pupils talk to the teacher and answer her questions just as though she were in the room. There is much less rote practice. Children are asking questions. Some schools are making barometers and undertaking other special projects. I saw one sixth-grade teacher bring in a lizard for the class to study. They talked about the little beast, what he ate, who ate him, and worked around to the balance of nature. This could not possibly have happened only a few years ago. I can't say that the high schools - which are using, mistakenly or not, very large classes - are working as well as the elementary. But on the whole, it is a very encouraging and hopeful experiment, one that is worth watching.

Pause briefly, before we review the next case, to reflect on what seem to be the basic elements of the Samoan case: a problem, large, urgent, important enough so that many groups can rally to solve it; adequate support, administrative and financial; and the introduction of television, not as a separate element, but as a part of a teaching-learning system integrating all the teachers and all the classroom activity.

CASE 2: HAGERSTOWN

Let us turn now from the fale huts and palm groves of Samoa to the green fields and active industries of Maryland. The United States experiment most often linked to Samoa is Hagerstown. Most of you are familiar with this system, and I shall not enumerate what you more than likely already know about it.

As you probably know, there were three chief reasons for introducing television into the Washington County school system, centering on Hagerstown. One was that school officials were anxious to make expert teaching of art, music, and science available throughout the elementary schools, and to offer certain advanced courses, such as college-level mathematics in high school. To do this with resident teachers would have added a great deal to the budget, and would have presented severe recruiting problems. Therefore, they wondered whether it was not possible to share expert teaching in these

subjects by means of television, at a lower cost, or at least no higher cost, than by hiring the necessary specialized teachers. In the second place, they wondered whether in the coming years of expanding enrollments there would not be some advantage in designing schools, not with many rooms about the same size, but rather with a central large room or rooms for viewing televised classes, and a number of small rooms around the perimeter of the building. And, finally, financial help was available – $1 million from the Ford Foundation and over $300,000 in equipment from manufacturers.

This was in 1956. Hagerstown decided, as did Samoa later, that it needed the equivalent of six channels to teach the core of an entire curriculum. Because open-circuit channels were not available in that number at any one point in the United States, it was decided to go closed-circuit. Coaxial cables were leased from the telephone company, and extended gradually over the next few years to every school in Washington County. For more than a half dozen years, now, every student in the system has received the core of his curriculum from television.

In Hagerstown, as in Samoa, a studio teacher and a classroom teacher operate as a team to teach content previously decided upon. The studio teachers are chosen from the regular teaching corps of Hagerstown, and are changed frequently. Because all the teachers in the system are qualified and well trained, there has been less difficulty, in the Hagerstown than in the Samoa experiment, in getting the classroom teacher to play his new role expertly. But two elements are generally credited with the smooth transition in Hagerstown from the older plan to the television teaching team. One of these is the firm support given the project by Superintendent William Brish. In a sense he played the role in Hagerstown that Governor Lee played in Samoa. The second is the insistence throughout on involving as many teachers as possible and thus reducing the potential threat of the new method. Only about 5 percent of all the Hagerstown television teaching is taped, and very little of it is brought in from outside. Thus teaching on television is treated as an ordinary function of team teaching, not specially rewarded, and passed around among promising candidates.

A considerable amount of research has been done at Hagerstown, and some of the results of introducing television have been spectacular. Remember, now, that this is not a deprived school system, nor was it in 1956. It is a rather average American system. But in Hagerstown junior high schools during the first four years of television the average performance of students on a standardized test of arithmetical concepts rose from the thirty-first to the eighty-fourth percentile, measured against national norms, and on a standardized test of problem solving it rose from the thirty-third to the sixty-eighth percentile. In other words, before television, two-thirds of the junior-high-school pupils of the country scored above Hagerstown children on these arithmetic tests; after television, two-thirds scored <u>below</u> Hagerstown. In the fifth grade, during the first year of television teaching, the average Hagerstown pupil gained 1.9 years, rather than the one year represented by national norms. In the year before television was introduced, every one of the grades three through six averaged below national norms in arithmetic. Within 2 years, each grade averaged <u>above</u> the national norms. We have spoken only of arithmetic. Not all results were as good as these, but the gains were by no means limited to arithmetic. Perhaps the most significant verdict was turned in by the school board of Washington County who, in 1960 when the Ford grant came to an end, studied the results, gave all teachers and administrators a chance to express their opinions anonymously, and then voted without a dissenting voice to take over the approximately $600,000 annual cost of the television in their own budget.

CASE 3: NIGER

Both Hagerstown and Samoa, one might say, profited by American know-how in television. Suppose now we turn to a country that has had little or no help from the United States in television. Niger has only sixty-six fully qualified teachers – meaning teachers who are secondary school graduates and have had teacher training. There are some hundreds more who have had the equivalent of junior high school education, and still more who have had only 5 or 6 years in school. But there is still no hope for adding to the resource in highly educated and trained teachers until the country's higher priority needs for educated people have been met. This creates an extremely difficult situation because Niger, which has had only between 5 and 10 percent of its school-age children in the classroom, is now trying to double and soon to triple its school population, and especially to strengthen its secondary schools. The problem was, then, how to find the necessary number of skilled teachers.

Acting with the advice and help of the French government, Niger developed a plan for using expert teachers in the studio and monitors in the classroom. Unlike Samoa and Hagerstown, which began with all the elementary and secondary grades at the same time, Niger began with the first grade. It began very cautiously, in fact, with only two schools, and with closed-circuit television. By the time the country's new open-circuit television station went on the air, the first-grade experiment had shaken down, and the television was expanded to the second grade. The plan is to add one grade a year.

Like Samoa and Hagerstown, Niger has found it works best to create in effect a teaching team in the studio and the classroom, and like Samoa it has found that the weakest link in the team is the classroom teacher when that individual is not adequately trained. Using elementary-school-trained monitors in the classroom, the supervisors of the Niger project have tried short in-service training courses, and have decided more intensive special training is needed. One of the ways Samoa tried to solve the problem of helping the relatively untrained classroom teacher to learn his new role was by putting into every school an American principal, who worked with his teachers and served really as a supervisor. They also brought all the teachers to a six-week summer workshop each year, and conducted daily in-service training by television. No easy way has been found anywhere to solve this particular problem, but it has reappeared again and again in developing

countries where instructional television has been introduced.

One of the distinctive features of the Niger experiment has been the continuing presence of a three-man research team supplied by France – a psychologist, a sociologist, and an ethnologist. These people are studying very closely the effect of television teaching on African children who have never before seen television, and, most of them, have never before spoken French, which is the language of instruction. Their psychological and cultural reports are not yet available, but they have measured the learning results of the first year. As might be expected, the results were better in some subjects than others, and somewhat better for children who were a year older than the others. But as a whole, the television classes averaged well above the average performance of non-television classes. We must say of Niger, as of Samoa, that all the evidence has not come in, but the outlook is highly encouraging.

SOME NOTES ON OTHER CASES

We will deal with several additional cases in a more brief manner.

You have noticed that in all three of the cases I have mentioned, generous financial help was available, and in the case of the two developing countries expert counsel was also freely offered. What would happen now if a developing country tried to go it on its own?

It isn't a recommended procedure, but it has been done. It was done, for example, in Arequipa, Peru. In that high Andean city, a group of school teachers, members of a religious discussion group, decided to undertake a community project. Looking around them, they saw that hundreds of children were unable to gain entrance into the schools because of shortage of places for them. Hundreds of others had left school after 3 or 4 years. They decided to begin with a group of the school-leavers who had gone to work as domestic servants. Suppose that the employers of these children would let them watch television an hour a day, they said. And suppose that the proprietor of a small local television station would give some free air time. And suppose that we were to do the teaching, and could get help in preparing learning materials. Well, all these things came true. Although with some objection, the employers let their domestic servants watch television; the station offered up to three hours free time; the teachers donated their time; the priest who was head of the discussion group supervised the curriculum. And a television school for school-leavers went on the air without a single paid employee. Since that time, the project has immensely expanded, now has a government contract and some paid employees, and offers a varied group of out-of-school courses, adult courses, and so forth. It has been almost fiercely local, resisting rather than inviting outside advice and help, and this has not been wholly to its advantage. But at least the Arequipa story shows that the do-it-yourself spirit is not entirely missing from instructional television in the new countries.

We have reviewed several uses of instructional television in school. The Arequipa project was mostly for out-of-school use. There have been a number of other projects of that kind, aimed at extending educational opportunities beyond the school – for instance, Telescuola which provides secondary education for Italian children who do not have secondary schools in their areas; and the Chicago Junior College which for nearly 10 years now has offered a complete junior college curriculum by television, with the home television students doing consistently better than classroom students on the same examinations. Television has been used to help teach literacy – in places like the Ivory Coast where it has helped to train native workers to read, write, and speak French, and do simple arithmetic so that they could be made middle-level supervisors; and in Italy where about 15,000 adults annually register to study reading, writing, and arithmetic, with the aid of the famous program series, "Never Too Late." The trend of the evidence in literacy experiments, though, is that, whereas television can assist greatly by providing visual material, by motivating a group, and bringing a certain amount of expertness, the real key to what happens lies in the ability of the group supervisor. A straw in the wind is that Italy, which at first used only monitors in its viewing groups, now uses certificated teachers – usually working outside school hours – for that purpose.

One more note: from almost everywhere instructional television has been used comes word that it is an excellent vehicle for in-service teacher training. The mere example of excellent teaching televised into a teacher's own subject is itself effective. As some observers in Samoa remarked, this may have been the first time that some of the native teachers had ever seen a well-made lesson in their own grade level and field. Furthermore, to prepare the teachers for teaching an unfamiliar subject or method, like the new math, television has proved an excellent in-service training tool. It was used for that purpose in Colombia, for example, and the results were carefully measured. The amount of learning was very high, but with teachers as with pupils the total amount depends greatly on what happens on the receiving end. That is to say, whereas a considerable amount was learned from television alone, a great deal more was learned if the teachers met in groups and discussed the broadcast lesson, and still more if the groups had competent supervisors to direct the discussion.

We have said enough now to indicate that instructional televison, used well, can contribute toward the solution of a great variety of educational problems – the need to raise instructional level quickly without a teaching corps adequate for the purpose, making up for the scarcity of specialized or highly trained teachers, extending the school beyond its physical boundaries, helping to teach adult literacy, and providing in-service training for teachers. This is not, by any means, to say that it always works well. I could cite some very disappointing stories to balance the encouraging ones I have mentioned. But I believe we know enough now to advance a set of propositions as to what are the conditions under which it is likely to work effectively. Let me suggest some of these.

CONDITIONS OF SUCCESS

(1) A problem.

After twenty-three case studies, I devoutly hope that the next time someone comes to me for advice on instructional television, he comes with a problem and not with a piece of technology. This is why some of the more spectacular examples of use of television for teaching have come in developing areas like Samoa and Niger, Colombia and Peru. In those cases the problems have been so challenging that it has been possible to think boldly of new ways of solving them, and to get sufficient support behind the effort. I think it is fair to say that most American school systems have not felt quite that order of need for television. It is likewise true that most American schools which have tried television have been the schools that needed it least. Perhaps not until we turn the power of instructional television on our deprived schools (which are to be found on the mainland as well as in Samoa and the Trust Territories), or to meet the likely expansion of schooling from ages 3 to 20, with thirteenth and fourteenth grades appearing in every city of over, say, 15,000 people, will we ourselves be challenged to use the medium to its full capacity.

(2) A teaching-learning system.

Instructional television hasn't worked well when it has been used like films - at the option and at the time of particular need felt by the teacher. Perhaps some day when we have enough inexpensive videotape recorders in the schools it may be feasible to use it that way. But the way it has worked most effectively, whether in developing or developed countries, is when it is integrated into a teaching-learning system - not when it is used alone, or when it is used as an optional visual aid. This requires the teacher in the classroom to learn a new role, and to many teachers at first this role may be threatening or uncomfortable. This has been the source of most teacher resistance, where it has occurred. But experience indicates that the role need be neither threatening nor unpleasant; in fact, it may be more satisfying than the role of being the only teacher in the classroom and in complete command of everything that happens there. And once the transition period has been passed, both teacher and student attitudes toward the new way of teaching improve steadily with experience with it. The experience of working in a cooperative team of studio teacher, classroom teacher, materials expert, and others as needed, can be a most satisfying one, once the roles are learned.

I would like to digress and suggest that one of the old stereotypes about television teaching - that it is a passive experience for the pupils - is destroyed in every place where instructional television is being used effectively. The classes are extremely active. The pupils talk to the teacher on the tube, practice as directed, answer the teacher's questions, pay as close attention, and, in elementary schools at least, seem to enjoy the experience at least as much as with face-to-face instruction. The television-teaching-team is not an encouragement to passivity, but rather an opportunity to stimulate learning activity.

(3) Support.

Any innovation needs support. Instructional television, which is expensive and somewhat disruptive to old roles, and which requires obedience to central schedules, needs it especially. Governor Lee in Samoa, who was strong and determined enough to meet the resistance he encountered from certain mainland teachers who did not believe in television; Superintendent Brish who was firm without being rough; the President and other chief officials of Niger who quieted some incipient uneasiness by themselves coming to visit the experiment - this kind of support is priceless in the infant years of instructional television. And beyond that, the financial support necessary to have an adequate technical service and an adequate nucleus of skilled personnel, is a sine qua non. The magnificent technical equipment of Samoa has made a difficult task much easier. On the other hand, I could tell you about small stations where the power is forever failing, where receiving sets are not maintained, where teachers are struggling with little help, little feedback, scant experience. I do not have to tell you that there is little chance of using ITV effectively under conditions like those just cited.

(4) Size.

There is a certain critical mass beyond which instructional television seems to make an important impact and works most efficiently. Part of this, of course, comes from the economies of scale which are made possible by larger operations. Colombia, for example, can deliver television for about 5 cents per pupil hour, but it is serving a quarter of a million pupils. Some unit costs, where the system is notably under-used, run as high as $2 per pupil hour. We find that in most systems, instructional television is under-used rather than over-used. Spectacular differences in unit costs can be achieved even by balancing up the investment in programs against the investment in receiving equipment. This is something we cannot develop in this paper, but you can find details in the case studies. Beyond the financial effect, though, there is a psychological impact that comes with size. It is impossible not to take the Samoa operation seriously. It is impossible not to take Hagerstown seriously. The teachers, the students, the communities know it. Observers have reported from several places that the very fact of the existence of a considerable ITV operation has made all developmental change easier in that area.

(5) Planning.

Let us merely say that most countries and school systems have bumbled into instructional television. Many of them have wakened and found it under the Christmas tree - the equipment being intended mainly for entertainment or persuasion rather than education. Many countries that have launched into television have been overwhelmed by the sheer logistics of getting on the air. There has been all too little time to sit back and decide what the new tool might best be used for, and what related changes might be made, and what will be required outside of television programs to make the new system work. And yet there will never be a better time, than just before the introduction of a powerful new teaching-

learning system like television, for a school system or a developing nation to sit back quietly for a while and consider goals and means and desirable directions of change.

There is no magic about instructional television. It is merely a pipeline into the classroom – a way to share and redistribute teaching resources. And yet it puts a wand of potential magic into the hands of any school system that has a real need to use it and the willingness to learn to use it well.

FOOTNOTE

1. The countries (or other political units) were Algiers, Australia, Colombia, Ghana, Honduras, India, Italy, Ivory Coast, Japan, New Zealand, Niger, Nigeria, Peru, Samoa, Thailand, Togo, United States, UNRWA. The case studies have now been published under the title, New educational media in action: case studies for planners. Paris: UNESCO and International Institute for Educational Planning, 1967, 3 vols. An interpretive volume has been published with them, Schramm, W., Coombs, P. H., Kahnert, F., and Lyle, J. The new media: memo to educational planners. Paris: UNESCO and International Institute for Educational Planning, 1967.

REFERENCE

Beeby, C. E. The quality of education in developing countries. Cambridge, Massachusetts: Harvard University Press, 1966.

Instructional Research: Some Aspects of Its Status,

Arthur A. Lumsdaine
University of Washington
Seattle, Washington

I WOULD LIKE to structure my paper in terms of the following main topics: (1) instructional-media research as related to research on learning and instruction more broadly; (2) classes of problems which are involved in the assessment of instructional effects, both at the applied level and at the "basic research" level; and (3) some considerations regarding the priorities of what it is most important to do in the instructional-research field at the present time. In relation to the question of priorities, I will also review the status of alleged "findings" of hopefully relevant basic research; some serious deficiencies of such "findings"; and some notions concerning requirements for improving the state of research practice.

PUTTING INSTRUCTIONAL-MEDIA RESEARCH IN CONTEXT

The first point I should like to make is one which people working in instructional-media research have realized in varying degrees for a long period of time. This is that the important problems in instructional-media research are actually broader problems of research on the effects of the management of learning and hence basically research on learning. Most of the laboratory research on learning has at least potential, even if unrealized, relevance to the design of instructional materials and procedures. However, much of it does not seem very useful when one tries to apply it to practical problems of instruction – as those who teach courses in the psychology of learning often realize when they attempt to apply their own principles.

Most of the important problems for research on basic variables in instruction--i.e., stimulus and response variables influencing the effectiveness with which instruction produces learning--have received some attention in the field of instructional-media research. In my opinion, in fact, more incisive research on instructional variables has, thus far, been done within the instructional-media field than in any other context. I think that this statement is not just a reflection of a parochial bias on my part; rather, I think it reflects the fact that the encapsulation of instructional procedures in specific media which are mechanically reproducible (like films and self-instructional programs) gives a degree of control over the variables involved in presentation, exposition, instruction, and persuasion that is very hard to achieve in other forms of research. Where a particular experimental treatment can be carefully and deliberately implemented in advance, as in a film in the cutting room, there is much less dilution from the concept, or variable, or procedure presumably being implemented, to its actual implementation in the process of instruction.

During the past few years, I have been especially concerned with what has been referred to as "programmed instruction." Like other catch phrases in the education and educational-research fields (such as "systems," "team teaching," "ungraded," etc), this is a term that I wish had never been popularized. It is not what you call something, but the way you do it that matters. Also, it has been insufficiently recognized that the work in the programmed-learning field needs to be considered as a special aspect of the broader field of instructional media (although some

of the Skinnerians don't like to think of it that way). "Instructional media," in turn, is part of the still more general field of instruction, which is a special aspect of the more general field of human learning. At each level it seems desirable to go back to basic learning principles where possible, as a source (but not the only source) of hypotheses or ideas for development and test.

Not all of the good suggestions for effective instruction come as safe deductions from Hull's postulates or any other simple, single source. Such suggestions also stem partly from analysis of subject matter--e.g., in Gagné's analysis of learning sets--and partly they arise in other ways. For example, in the process of developing specific instructional products one learns a great deal about what variables seem to be important or worth looking into. One formulates hypotheses that a learning theorist, as such, would never think of when one starts to attack specific, concrete problems in education. A related point which has been made in other contexts is that it is generally unwise, in my view, for people who wish to do fruitful basic research in learning to get too far away from the actual problems of school subject matter, from the things that learners really have to learn. On the other hand, I also think that it is important, in trying to analyze the essentials of instructional-media (or other instructional) situations, to have at least a good talking vocabulary with which to interact with people whose main concern is the basic psychology of learning.

I have stressed that the instructional-media field, in its most important aspects, must relate closely to the more general field of research on instruction and learning. However, I also emphasize that there are certain features of instructional media that make them very useful as tools for research, and also objects of research (cf. Lumsdaine, 1963). Thus, the general property of stimulus reproducibility in instructional media greatly facilitates scientific study of instructional variables. A related factor is that in producing instructional-media programs the implementation of instruction is in the hands of a relatively small number of people, thereby allowing less dilution in the application of instructional principles, than when one tries to implement these same principles by influencing the behavior of thousands of classroom teachers. Finally, the ability to reproduce instructional materials widely means that application of applied research results is multiplied many fold. It is more reasonable to put out a large investment to improve the effectiveness of materials which will be used by millions of students than it is to invest comparable resources in research on procedures which are only going to be used by a much smaller number of students.

SOME GENERAL QUESTIONS FOR RESEARCH ON THE EFFECTS OF INSTRUCTION

I would like now to identify some of the questions which I think one should ask about the effects of instruction and which I think are logically prior to the formulation of specific research studies. The first question is, "What <u>should</u> the effects of instruction be?" Second, simply, "What <u>are</u> they, in any given instance?" Third, "What <u>accounts for</u> them?" (This is where a great deal of research emphasis has been.) Fourth, "How do you <u>predict</u> them?" Fifth, "How do you <u>improve</u> them?" And sixth, "What is the <u>cost</u> (and 'cost-effectiveness') of improving them--how much is worth paying for a given amount of improvement?"

WHAT EFFECTS SHOULD BE

A number of people have pointed out that in evaluating instructional programs there are several kinds of criteria used. In the report which was published by the Joint Committee on Programmed Instruction and Teaching Machines (1966) three classes of criteria were distinguished: first, <u>appropriateness</u> of various possible effects; second, the <u>effectiveness</u> with which sought-for outcomes were brought about; and third, the <u>feasibility</u> of using the materials, including cost. Considering just the first two, it is quite evident that one can do a very effective job of teaching the wrong thing. (It could even be held that doing an <u>ineffective</u> job of teaching the <u>wrong</u> thing may be better.) What I want to call attention to here is the far underdeveloped research field of constructive and imaginative attempts to improve on the basis for determining what <u>ought</u> to be taught. Very little effort has been put forth on this as compared with approaches on how to teach a given thing more effectively. This applies all the way from kindergarten to graduate work.

Skipping over the second question, consider Question Three: What <u>accounts for</u> the effects of instruction? Here we have perhaps attained the stage of sophistication in which we turn away from unanswerable questions (such as, Which are better, films or lectures? or, Is conventional instruction better than television?), and have moved to questions of what variables account for particular effects. That is, we have become better aware of the consequences of the obvious fact that a good film will always beat a poor lecture and vice versa, and of the fact that such terms as "conventional instruction" have been so poorly defined as to be almost completely ungeneralizable.

Question Three gets at the attempt to manipulate variables in order to see which kinds of stimulus-input variables make a difference in instructional outcomes. Obviously, if this type of research were to be successful in building an applicable science of learning, within the confines of a given set of instructional objectives, we would then be able to better answer Question Four, (How do you predict the effects of instruction?) as well as Question Five, (How do you improve them?). That is, we would have a basis for evaluating an instructional program of any sort by looking at it, analyzing its properties, seeing whether it conformed to characteristics of instruction which have been demonstrated to be more effective than other forms of instruction, and thus would be able to evaluate, <u>a priori</u>, any particular piece of instructional material, or program, as to its effectiveness.

Some of my colleagues feel that this can now be done with respect to programmed instruction. In my opinion, they are quite mistaken. For example, in one of Rothkopf's (1963) studies, (admittedly one

which is too limited to be definitive) a group of "experts" (students who had recently been through a course in the "proper" way to construct instructional programs) were asked to rate seven programs. These seven programs all dealt with the same subject matter and, without the knowledge of the judges, their actual effectiveness in teaching specified kinds of outcomes had been previously determined by trying each of the programs on seven randomly selected groups of students. When rated effectiveness in that set of seven programs was correlated with measured effectiveness (as measured by what students actually learned), Rothkopf concluded that the instructional effects were indeed "predictable": the correlation was minus .75. All that need be done was to get judges to rate them and then do exactly the opposite. Such data support the contention that the effects of instruction are not really very well predictable on the basis merely of inspecting the features of the instructional materials and deciding whether they conform to the implications of a "science of learning" or to the accumulated wisdom of teachers. This means that either we are applying the wrong findings or we have a great deal more to find out about the effective variables in instruction. This in turn takes us back to Question Two, What are the effects of instruction?

I contend that we might well set for ourselves the perhaps modest objective of finding out what we are actually doing in much on-going instruction. I think we at least need to know what effects we are producing before putting such exclusive concentration of effort into the more elegant question of what specific variables account for these effects. In many of the studies of instructional effects, one problem has been the lack of significant differences, reported with dismal uniformity. This lack would have been predictable if the investigators had bothered to run an initial test of any one treatment of the instructional materials (preferably the best), using available criterion measures of output. Often this would have shown them that, in fact, the whole program was teaching at such a low level of efficiency that the likelihood of revealing any differences among the treatments was indeed very poor. All one has to do is to read the instructional research literature to find many instances of this kind, and I must even admit that this has been true of some of the studies which I have been responsible for sponsoring. It is a very simple truth, but one which has eluded investigators for a long time, that if we are going to investigate factors that make instruction effective we have to get the instruction effective in the first place and then find out something about how it got that way, instead of working hard at manipulating variables in a piece of instruction which we later find out is operating at a low level of efficiency. One of the implications of this is that the classical sequence of research and (then) development often must be reversed. In other words, one must do the development first; only when something effective has been developed, do we then have an adequate tool with which to conduct useful research. This is an oversimplification, of course, but it remains a crucial and an all-too-often ignored consideration in the instructional research currently being done (and continuing to be supported).

Regarding Questions Four and Five (How to predict the effects of instruction and how to improve them), it may be noted that both imply important kinds of purposes. However, it can be argued that prediction is of no use in itself unless it can ultimately be used to bring about better instruction. One way to do this is through better selection of existing materials (as well as making better ones).

The conduct of adequate empirical evaluation studies for this purpose, which would require the detailed testing of every piece of marketed instructional material, is an extremely time-consuming business. It would obviously be a great advance if one could reliably predict which materials would be more effective than others, as a basis for selecting from available materials within a given set of instructional purposes. In any case, the ultimate applied purpose of fundamental knowledge about the influence of instructional variables on learning outcomes is to be able to know how to improve the instructional product.

A vexing problem of instructional research for this purpose concerns the extent to which the independent variables named by the investigator are idiosyncratic to him. For example, in experiments on "student participation" or "varied repetition," (or you name it), the independent variable is seldom defined in terms sufficiently operational to provide reasonable assurance that someone else would be doing essentially the same thing if he said he were studying the nominally "same variable." That is, for most of the pedagogical variables derived from either common teaching practice or learning theory, the specific way in which the "variable" is implemented can turn out to be a great deal more important than what variable it nominally is. We, therefore, have findings in which the "same variable" (e.g., overtness of response) turns out to have one effect when explored by one investigator and another effect when explored by another investigator. The conclusion would seem to be that, pending such time as we are better able to define what our instructional variables mean in reproducible, operational terms, we had better assume the burden of demonstrating explicitly the generality of our findings. The deduction from this is that in an instructional experiment in which one manipulates a named variable or procedure, the critical number of "degrees of freedom" on which generalization forms the experiment depends is a function of the number of experimenters (or programmers), rather than the number of subjects (learners). Put another way: if we are really going to say that we have found a dependable rule or generalization about an instructional procedure, we have got to show that we can write a set of instructions which will allow another experimenter, programmer, or teacher to create substantially similar effects to those we obtained. I know of only one or two studies (still underway) in which this is systematically being done. In one of these studies the number of instructional vehicles representing each treatment are written by six different programmers following a given instruction, as against six other programmers who are following a different instruction. The number of degrees of freedom in that analysis will be something in the order of either 10, 12, 5, or 6, depending upon what kind of statistical test is used, but it will not be a

function of the number of students or classes that are taught.

This very important point can be summarized simply by saying that we have to be able to show, in a scientifically acceptable fashion, that someone else can manipulate and implement the variable under consideration. At present, I know of no way to do this, for most variables at least, except by brute empirical demonstration. This is one of my main reasons for the quotes around the word "findings"; for, of the experimental results on instruction, very few, if any, have been systematically replicated in a way that provides this kind of formal demonstration of their generalizability outside of the particular program or programmer responsible.

To consider Question Six (concerning cost and cost effectiveness) requires that we first consider basic objectives. In terms of the kinds of outcomes we are trying to achieve, we have ultimately got to translate progress toward these objectives into something that is cost accountable. I think the things that have the best chance of becoming cost accountable are things that are correlates of time, rather than being correlates of achievement level. For this reason and other, I think a strong case can be made for Skinner's contention that, in effect, we should use as our dependent variable the time to achieve a given minimum criterion of proficiency, rather than the levels of proficiency attained after a fixed number of hours of instruction as shown by the resulting distribution of scores.

A related problem is that our realistic goals in education are often how to bring people up to a point where they can quickly relearn rather than to maximize how much is retained by the end of the semester. There is simply getting to be just too much to know and too much to learn to keep it all on tap, and part of the problem of education is, increasingly, going to be to determine how much we ought to get "kept" in a man's head versus how much we ought to help him out on the job by better information sources or refresher training. We might ask how much a realistic criterion of the goodness of instruction is the speed with which a person can get back up to proficiency, rather than how long it takes him to get there in the first place.

A difficulty with this kind of criterion (aside from the break with tradition that it implies) is that it greatly increases the cost and difficulty of experimentation. If one is going to measure relearning of paired associates, the savings method of the psychological laboratories is a rather easy method to apply. But if one is concerned with the learning of a subject matter like calculus, instead of pairs of nonsense syllables, then the question arises: What is the material from which one does the relearning? I think we have to work hard on solving such problems, because time expended in order to attain or regain acceptable proficiency offers the best promise of being translatable into cost terms. We can form some estimate of the cost of the student time and can likewise estimate the cost of the heating for the school, the teacher's time, etc. But how to estimate the cost of the difference between 700 and 750 on a set of achievement test norms seems much more puzzling, and possibly involves the wrong kind of educational objective anyhow.

RESEARCH AND DEVELOPMENT PRIORITIES

QUANTITATIVE PRODUCT-EVALUATION DATA

I believe that a great deal more work is needed on the simple, modest question of assessing or measuring the effects of available and reproducible instructional programs; I think this perhaps unglamorous area rates a very much higher priority than it is getting. We need to know what we are accomplishing, and by and large we don't know that. The programmed-instruction people have banged the drum for data accompanying programs, and we have even set up some standards so that we could have some hope that a sort of Gresham's Law wouldn't operate with bad data driving out good data. Yet at present I know of only a very few programs on which any reasonably adequate data have been published. Somewhat more than that number of programs have something which purports to be data, but which are merely records of what a few students did as they went through some version (not necessarily the published one) of the self-instructional program. For the present I would allocate a much larger share of research budget to support lowly, "dirty," applied work on getting product specifications for available programs that potentially will have wide use, and at least relatively less of it to what are, all too often I think, futile attempts to do fundamental research. (I am well aware that some of my colleagues will regard this as utter heresy.)

THE PROBLEM OF WHAT TO TEACH

Another act of high research priorities is related to Question One, – What should the effects of instruction be? We have an enormous range of opinion on this. We also have an increasing amount of knowledge to be taught and yet no rational basis for selecting what we should teach, not only to graduate students in specialized areas and to undergraduates who are acquiring a "liberal" education (presumably to prepare them in some fashion to deal with the problems of the world), but even, to a considerable extent, to the elementary-school children. The problem does get somewhat easier at the lower grade levels, as there are certain fundamental skills which appear necessary; but beyond the elementary-school grades I think we are flying blind in terms of what we are teaching. A very good case can be made, for example, for teaching high-school students the fundamentals of sampling and scientific inference rather than Euclidean geometry.

In any case, I believe it would be very helpful to try to do more extensive survey research to attempt to identify areas where agreement and consensus can be reached as to what should be taught. At this time a very promising attempt of this kind is being made by a group under the leadership of Ralph Tyler (1966).

I also feel that some experimental evidence could be quite useful in sharpening up the question of what to teach. Many of the questions of what ought to be

taught are ultimately transfer-of-training questions. If a more sustained attempt were made to define and characterize the kinds of educational outcomes we really want, then we could try to verify some of the working assumptions which we seem to make as to what kinds of learning transfer to these outcomes. We would be then moving in the direction of an experimental attack on what ought to be taught. There is, of course, so much hypocrisy in the whole field anyhow that even to get people to be a little more honest would be helpful. ("The purpose of my course is to teach the students to think"; but the final examination consists of multiple-choice questions about the facts in the subject matter.) Very few determined attempts have been made to formulate instructional course objectives so as to reflect in the final examination, as well as in the development of the course, what the professor professes is really the most important objective of his educational endeavor. Some hopeful steps in this direction have been taken. (The last course I taught got almost to first base in this respect, and the next time I teach it I hope I can get a little further along.)

EMPIRICALLY GUIDED PRODUCT DEVELOPMENT

Next to a determined attempt at research on what should be taught, I would say the next most important priority within the total research and development sphere is not in the research field proper but in the area of development. In my opinion, a great deal more support than is presently the case should be devoted to empirically guided development projects. (The U. S. Office of Education has not as yet fully taken my advice on this point.) Perhaps one way to proceed would be to require product development as one element of an R and D package. That is to say, if one is to conduct an experiment which purports to "test a hypothesis" and derive some "fundamental knowledge," and if one is realistically pessimistic about the likelihood that he will succeed in this objective, perhaps he might be asked to have a second string to his bow. If, for example, all such experiments were conducted with actual subject matters taught in schools and did both of these two things, then if they failed in what was the nearest and dearest objective to the heart of the investigator, they still would have accomplished something useful. That is, if one produced a useful package of demonstrably reasonably effective new instructional materials that were needed as a vehicle for the experimentation, and if this product of the research were to be made available to some appropriate distribution agency, then the sole product would not be confined to a dubious conclusion in the pages of a journal; the product would include an actual on-the-shelf instructional item that could be used. If in addition a little illumination was shed on the basic dynamics of instructional variables, so much the better. If it weren't, at least one would have a useful film, self-instructional program, book, etc. which would help to justify the investment that went into the research project.

Also, some studies could usefully take an existing product, widely available and in current use, and use it in a research project as the vehicle for making experimental manipulations to inquire about the effects of instructional variables. Then, data giving a good profile on the actual teaching accomplishments of this product could be made available in the form of product-specification data. This would thus provide a well-defined set of measures, analyzed item by item, so that one would know which points of information, skill, knowledge, or other behavioral outcome were increased and by how much by the use of the basic product in a specified manner with specified kinds of learners. One would then at least have some useful product data, regardless of whether or not his attempt at scientific inquiry on instructional variables were successful.

In summary, I would tend to plug first for empirically-guided development studies, second for other evaluative studies, and (a low third) for purely "basic" research, given the present state of the art. A good way to have one's cake and eat it too here is to do studies which combine investigation of instructional dynamics with the development or assessment of a product, in such a way as to create either a useful product or useful product data, in addition to whatever illumination may be shed on more fundamental, generalizable problems.

SOME WEAKNESSES IN "BASIC" RESEARCH "FINDINGS"

As to the frequency or likelihood of genuine illumination, suppose that one applies the criteria any real scientist has the right and obligation to apply to research findings, including that of replication; and that one throws out all of the findings which lack any adequate guarantee of the sensitivity of measuring instruments or fail to show sophistication about power of statistical tests used, and that one throws out all of those indefensible findings which have proven the impossible by violating the injunction against proving the null hypothesis. One then gets down to the relatively few findings in which significant differences are reported, of which he will still find that very few have been replicated. Of the small number of these, perhaps most are in the field of the role of active response of students. Even here, however, it is difficult to say much more than simply that some forms of student responding under certain circumstances can prove quite helpful; the conditions under which one form or another of responding can help how much are still far from clear. Also, there seems to be some pretty consistent evidence that we have probably underestimated the extent to which we ought to help the student out by prompting; but that evidence is more secure for paired-associate learning than for more involved kinds of instructional materials.

A more gross generalization that I think could be supported as a kind of by-product is that most instructional lessons, films, television lessons, and college courses try to teach more than they possibly can in the time that is allotted them, at least on the frequent assumption that these are the sole medium of instruction. Demonstrably, almost every instructional program as initially presented is found to need to be extended, to need to be given more time (or else its objectives need to be reduced to more modest size) when the program is tested to reveal a profile of what students actually learn. This can be

found in study after study when the results have been reported in such a way that you can find out what was actually learned, not just in terms of P-value or a misleading dichotomy between "significant" and nonsignificant findings.

The conclusion may be a discouraging one, but the most important variable in the effectiveness of instruction all too often seems to be simply the amount of time which is spent in instruction. Thus, if one wishes to establish the importance of a procedure one has to do a lot more than has customarily been done in ruling out not only the null hypothesis but also what some of us have come jokingly to call the "erg hypothesis." Too often the procedures which are shown to be better just happen to be also the procedures that take more time, and with no sensible way of refuting the skeptic who says "So what? Naturally, if you spend more time you will teach more."

Many of the weaknesses of the instructional research to date are, I think, due to the lack of the promulgation and enforcement, for so-called basic research studies, of adequate standards of scientific reporting (which we have tried to formulate for evaluation studies in the Joint Committee report mentioned earlier). The generally poor standards of reporting allow the ambiguous use of terms, and failures to distinguish interpretations from actual findings, to a shocking degree in much of the educational research literature. It is almost impossible to read through a set of papers in this literature without finding, upon examination, that many conclusions stated as if they follow from the data are really just some speculations produced by the experimenter to explain what he found or why he didn't get what he expected to find. I am exaggerating only slightly. It is an agonizingly difficult task to attempt (e.g., Lumsdaine and May, 1965) to glean from research literature in the instructional-media field that it is that has actually been demonstrated. In many instances one is unable to get the basic data necessary to find out whether the findings warrant reporting, yet perusual of the available reports makes it clear that without going back to the basic data one cannot determine what conclusions, if any, the available data justify. One simply cannot believe, by and large, what one reads in abstracts, secondary sources or even in primary research reports. And if one tries to check on the latter, he often finds that the basic data are inaccessible.

The APA-AERA Joint Committee recommended, for evaluation studies, that the basic data should be preserved, preferably in the form of a complete matrix of each student's response to each test item, so that the summary results can be verified. The value of such data preservation is twofold: (a) very often serious errors are made in the data summarization; there are nothing like the uniform standards of checking in educational research data that one expects in physics or in the experimental psychology laboratory; and (b) very often the findings, even if accurate, are recorded in a summary form that precludes answering relevant additional questions to which the original report was not addressed. By way of example, there is a very good a priori reason and considerable empirical support for believing that the effects of different kinds of instruction are likely to interact significantly with intellectual ability, as measured at least grossly by IQ, Army General Classification Test scores, etc. Yet study after study appears in the literature in which only the overall data are reported, with no indication whatever, even though the basic data contains this information, of how the results interacted with, or were modified by, differences in student ability Even when differential analyses are a function of ability are made it is often impossible, without going back to the basic data, for one study to be compared with another which has, let us say, made one split on an ability measure while another study has made a different split, or used a different measure. It would be perfectly possible, and often of considerable interest, to see if these studies indeed contradict or confirm each other, but a person cannot do it because he doesn't have the same split on the differentiating variable. This often would be very easy to rectify if the basic data were preserved, as also would be true in other instances in which the analysis has been incorrectly performed but might be corrected by going back to the basic data.

Clearly I believe there is great need for general improvement in reporting standards. As another example, one of the most common errors that we try to teach introductory students to avoid is that special form of the unprovability of the null hypothesis which is exemplified in the following argument: Group A and B did not differ significantly before getting treatment X; they differed significantly after the treatment, and therefore the effect of the treatment was significantly different for the two groups. Of course this argument is quite invalid; what can be shown rigourously is only that the difference between before and after was greater for one group than for another. It is quite possible to document the proposition that treatment X made more difference for group A than for group B but, again, the basic data by which this might be tested (e.g. by a second-order t-test) have not been preserved so as to be available for making the test.

SUMMARY

In summary, I suspect that the most important thing that can be done to improve instructional research methodologically is simply to try to clean up the standards of reporting and thereby the standards of performance for educational experiments--at the semantic level, at the data-preservation level, and at the level of insisting on a more complete description of what was actually done in the experiment. If this is not done, and here I utter a Cassandra-like note, I believe ultimately we are going to kill the goose that is laying all the Research and Development Center eggs. We need desparately to clean up our experiments methodologically, and to insist on at least the same standards of rigor for educational research reports and journal articles that are applied by the editors of the Journal of Experimental Psychology.

I should like to add that the dependent-variable measures we are using in educational experiments are also in great need of improvement. For instance, there are few studies available in which the effects of instruction or indoctrination on "attitudes" have

transcended the level of mere verbal statements on a questionnaire by the respondents as to how they think they were affected or how they believe they would behave in a particular situation. Only a few studies have gone beyond this. One such study was conducted by Mark May at Yale some years ago (May and Jenkinson, 1958). Instead of simply asking the students, May observed how students actually behaved so as to obtain a measure of motivational influence of two alternative forms of a film. George Gropper and I did a somewhat similar study a few years ago in which we assessed the effects of some television broadcasts in alternative forms presented over WQED in Pittsburgh (Gropper and Lumsdaine, 1961). Janis and his associates at Yale have also used observation of real criterion behaviors, instead of mere verbal surrogates, in studying the effects of various forms of communications (see Janis and Feshback, 1953).

But such studies represent only a bare start in the direction of what needs to be done. That this is so leads to a final conclusion I would like to draw: namely, I think it would be much better if we would pick out a relatively small number of problems, and a relatively small number of standard subject matters, for such experimentation as purports to get into the real dynamics of instructional variables, and stick to them and really try to pin a few things down by extensive varied replications with good dependent measures, instead of striving so much after novelty that we hop from problem to problem with a series of superficial studies that cannot be compared with each other and that really prove nothing, individually as well as collectively. I believe this can be said, as a matter of fact, throughout the whole spectrum of educational-psychological research. There are those who advocate, and a few who even put into practice, the insistence that Master's theses should always be replications of experiments rather than attempts at new experiments. I personally think that this would be a healthy general practice; for instructional research I know it would be. We need a great deal more replication and elaboration of a few putative findings that look fairly promising, and much less uncoordinated striving after novelty in an attempt to win kudos in "pioneering."

REFERENCES

Gropper, G. L., and Lumsdaine, A. A. The use of student response to improve televised instruction: An overview. Report No. 7, Studies in televised instruction. Report No. AIR-C13-61-FR-245 (VII). Pittsburgh: American Institute for Research, 1961.

Janis, J. L. and Feshback, S. Effects of fear-arousing communications. Journal of Abnormal and Social Psychology, 1953, 48, 78-92.

Joint Committee on Programmed Instruction and Teaching Machines. Recommendations for reporting the effectiveness of programmed instructional materials. Washington D. C.: National Education Association, 1966.

Lumsdaine, A. A. Instruments and media of instruction. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963. Pp. 583-682.

Lumsdaine, A. A. and May, M. A. Mass communication and educational media. Annual Review of Psychology, 1965, 16, 475-534.

May, M. A., and Jenkinson, Nelle L. The reading of a book based on a film. In M. A. May and A. A. Lumsdaine, Learning from films. New Haven, Connecticut: Yale University Press, 1958. Pp. 195-203.

Rothkopf, E. Z. Some observations on predicting instructional effectiveness by simple inspection. Journal of Programmed Instruction, 1963, 2, 19-20.

Tyler, R. W. Let's clear the air on assessing education. Nation's Schools, 1966, 77(2), 68-70.

Defects, and Needs — SOCRATES, A Computer-Based Instructional System in Theory and Research

Lawrence M. Stolurow
Harvard Computing Center
Cambridge, Massachusetts

THIS PAPER describes a cybernetic computer-based instructional system, SOCRATES, the teaching model which led to its development, and some of the research accomplished with it. Early work on non-cybernetic instructional systems is described in Coulson (1962), Silberman and Coulson (1962), and Bushnell (1962). For a description of other existing computer assisted instructional (CAI) systems see Stolurow and Davis (1965), Gentile (1965), Hansen (1966), and Zinn (1965).

SOCRATES

The acronym, SOCRATES, is System for Organizing Content to Review And Teach Educational Subjects. It consists of a group of student input-output (I/O) stations wired to a digital computer through a relay rack. It is a computer-based, multiple access, instructional system designed to meet the requirements of an idiographic contingency model of the instructional process (Stolurow, 1961; 1965b, 1965c). The model defines the presumably critical categories of instruction and treats these as system functions. The system permits controlled implementation of the model as well as the collection of data which could lead to its further development.

Figure 1 shows the configuration of the system, called SOCRATES II. It consisted of an IBM 1620 computer with a 1311 disk for additional storage, plus a 1710 Multiplex and terminal unit linked to several student I/O stations, or consoles.

Figure 2 shows some of the individual student I/O stations, each of which displayed 35 mm frames of a film strip with a 1,500 frame capacity and provided the learner with fifteen keys for either single (selected) or multiple (constructed) response to a display (Stolurow, 1966).

AN IDIOGRAPHIC CONTINGENCY MODEL OF TEACHING

The model of the teaching-learning process is an idiographic contingency conception (Stolurow, 1965a, 1965b). It is based on the assumption that individualized instruction is a multiple stage decision process that is diagnostic, prescriptive and corrective in its use of psychological information to make decisions about teaching. While the decision process is basically continuous three stages can be identified. In the first stage the pre-tutorial decisions are made; in the second, the tutorial decisions are made; and in the third, the program revisions are made. In the first two stages there are two levels of operation which make the system highly adaptive to individual differences.

Pre-tutorial Decisions

In the pre-tutorial stage, one set of decisions determines the assignment of each student to an instructional station; the other set determines the educational program to be used in teaching a student.

Automatic assignment. When a student entered the instructional area, he could sit at any vacant student station each of which contained a MASTER I/O unit. The MASTER I/O unit, connected to the computer, contained a library of materials on film

Figure 1. SOCRATES II

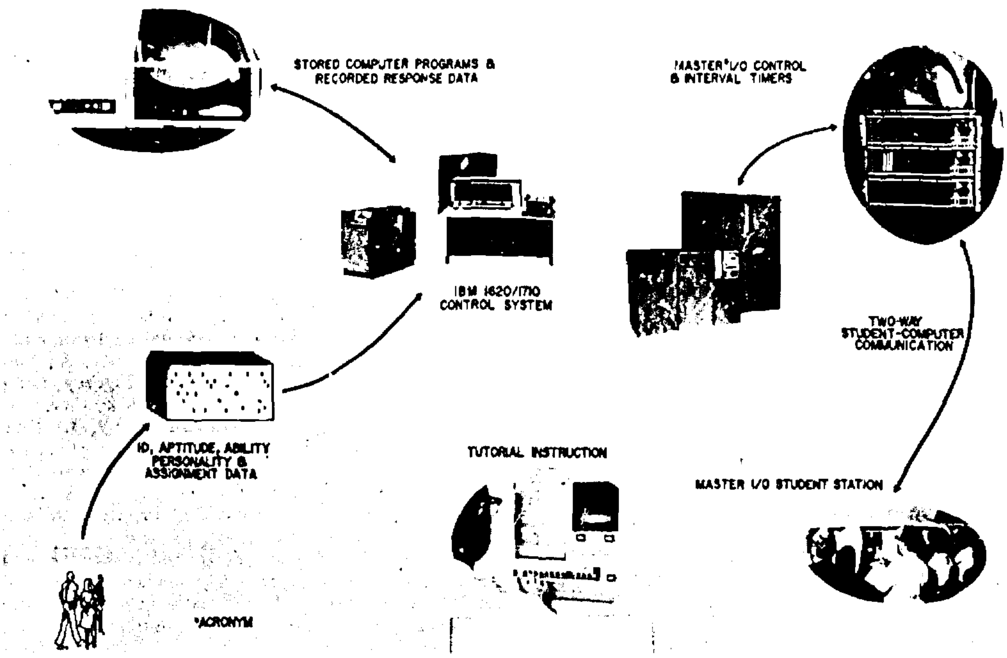


Figure 2. Individual student I/O stations

(1,500 frames) and each could contain a different library (film strip) if this seemed feasible and desirable. Feasibility depends, in part, upon the amount of core storage one needs for the computer programs used to process each of the educational programs contained in a library.

The purpose of the automatic assignment was to simplify and expedite the process of getting the student on-line with the computer. The procedure in automatic assignment consisted of the following. A student went to any station - for example, 11. If it was not one at which the instructional materials needed were contained, he had to be directed to another station which had the appropriate material. In order for the system to do this automatically, the student read an initial frame containing instructions to both enter his ID number using the keyboard and to press the "I am thru" key. When he did this, another frame appeared on which a film number was displayed. Every film contained a unique number. After the student entered his ID, he entered the film number. The system then compared the number with the one it had stored as the appropriate assignment for that student. If there was a match, instruction began at that station. If they did not match, however, the system searched all available stations containing the film which the student should see and advised him of the one available; it instructed him to go there. This was done by showing him a station number. Once the student was seated at an appropriate station, the pre-tutorial phase of instruction began.

Initial selection of an individual's program. In order for the system to make pre-tutorial decisions about the student's instructional program, it is necessary to have a computer program which processes information about the educational program and about the student. The objectives of the instructional program must be stated in behavioral terms, namely, the type and level of performance that is required. Each objective is a "do statement"; it contains a relevant verb and a description of content and conditions under which performance is to occur. Each film included frames containing: (a) information; (b) directions (e.g., procedure); (c) questions; (d) response options; (e) knowledge of results; and (f) evaluative feedback. Performance requirement for correct response, expected wrong responses, and the maximum training time that can be used are all specified. The student's pretest scores indicate his beginning knowledge in relation to each objective, his aptitudes, and his personality characteristics.

More specifically, whenever any of the student's pretest scores - his entry level of performance - equals or exceeds the minimum terminal level, the set of frames relating to that objective is eliminated for that student. The remaining set of objectives has to be taught; to do this, the system needs to be programmed to make the initial decisions about the rules to be used in teaching the remaining set of objectives. The set of decisions used in teaching is its instructional logic. When the logic is expressed as a set of rules for interaction with a student or, in other words, when the contingencies to be used with a student are decided upon, then the teaching strategy is defined and the program specified. The initial decision is to identify the instructional logic. The flow diagram for program selection during the pre-tutorial process is presented in Stolurow (1967).

Selection of contingencies. To use data about the student in making the decision about how to teach him, it is necessary to have information about the typical, or expected, relationship between pretest scores, on the one hand, and the effects of specific treatments such as a type of feedback or a sequence, on the other hand. Each of these relationships is a possible contingency that has to be implemented by the system. The desired contingencies are expressed as "if...then" statements. To illustrate: A contingency, as a rule to be used during learning, states that if a student with low math aptitude answers correctly, then he is sent on to the next frame, but if he answers incorrectly, then he is sent back to the original material and must select another alternative. This is a correction contingency. Another example is the aptitude-sequence contingency, namely, if the student's aptitude scores are X, then the sequence of instructional material is 1, 2, 3. For example, if a student with high language aptitude correctly answers the first five questions about a passage just read, then the remaining three questions in the set are omitted and he is given new materials that he would not otherwise see. However, if he does not answer correctly, the remaining questions in that set are given. The overall instructional logic is what determines the order in which the student would receive new materials during learning. If research data show that students above a particular level of mathematical aptitude learn to add fractions better through the RULEG sequence than they do through the EGRUL sequence (Taber, et al., 1965), then a contingency between aptitude level and the RULEG sequence would be programmed and used in making decisions about teaching the addition of fractions. Students with scores outside that range might be taught with another sequencing rule, e.g., asynchrony (Detambel and Stolurow, 1956), or class-descriptive cueing (Wulff and Stolurow, 1957).

To specify another contingency each student's personality test scores can be used to determine the form of evaluative feedback informing him of the correctness of his response (Frase, 1963; Parisi, 1965). These are just two examples of other contingencies that can be used in making teaching decisions. Ideally each contingency should be based upon data from research revealing that it contributes to the effectiveness of instruction. However, if appropriate data are not available, a set of theoretical possibilities can be used to specify the type of feedback, the frame sequence, etc. Whenever the teaching system uses contingencies, it branches the course of instruction at the decision points. A system can make a decision for each student at any point in teaching and alternative contingencies can be compared in their effectiveness to determine the value of branching. In doing this, the system is being used to test hypotheses about contingencies. Thus research needs have to be satisfied as well as those of application. All rules specifying the initial set of contingencies have the status of hypotheses and, as they are used, the system acquires the necessary data relating to their utility. This also means that the system provides a means for identifying additional contingencies. The instructional logic, expressed as a set of rules, is the set of contingencies used in teaching a student

using a particular set of materials. In individualized instruction the logic may be quite different for different students. The system must be programmed to select a strategy to go with the topics to be taught each student. The combination of content topics and strategy makes up an instructional program for a student.

Automatic adjustment. The system must also be able to handle the student for whom no progress could be expected to produce the minimum level of performance specified by the instructor.

Whenever the student's data indicate that he would probably not reach the instructional objectives by any available program, the system can a) teach the student some things to remove his deficiencies; b) lower its minimum acceptance criteria for knowledge; c) lower the terminal objectives; or d) reject the student. Each of these alternatives can be programmed so that it is considered automatically in terms of a predesignated priority rule. For instance, if the decision is to improve the student's proficiency, then he would be given remedial instruction in one or more of the prerequisites. For example, the student could be assigned and given an algebra program to build up his entry behavior level before he begins studying a statistics program. When he completed the algebra program, he would be tested and his new entry level of performance would be substituted for his original level and another decision would be made about teaching him.

Another alternative is to change the minimum acceptance criteria. The student below minimum-performance level on the prerequisite knowledge test but high in relevant aptitudes can be given an instructional program which is most likely to increase his final performance on the test of terminal behavior. Doing this would increase the proportion of failures, but it also would increase the absolute number of successes. Therefore, this is a decision strategy that probably would be used only if it was very important to meet a quota of trained people.

As a third alternative, the system can modify terminal objectives when it fails to find a suitable program. One way to modify objectives is to increase the maximum allowable instructional time. For each change in that parameter, the system can re-estimate a student's final performance to see if the next estimate meets the minimum terminal proficiency level. Once the estimated score equals or exceeds the minimum, the system can print this out so that the instructor can decide what to do next or it can compare the estimated time value with a new maximum value. The result of the comparison determines whether a program will be used with the student. A second way to modify objectives is to eliminate topics that have low priority. A third way is to lower the minimum level of performance which is acceptable at the end of instruction. Each of these ways of modifying objectives can be programmed so that the appropriate parameters are considered sequentially and a new estimate of final performance obtained. The system can store the estimates from each of these modifications and select that one which provided the maximum final performance estimate.

The system can be programmed to operate at two different levels in making pre-tutorial decisions. At the lower level, it makes decisions based upon a fixed set of rules and a fixed set of values for the critical parameters used by those rules. At the higher level, the system changes the values of parameters according to a different type of rule, one which determines the priority of the parametric changes and then re-evaluates the student. The priority rule itself can be varied from student to student to achieve further individualization. For example, the priority rule can be made to depend upon the relative size of the discrepancies between the original criterion values and each student's scores.

The outcome of the pre-tutorial processing is a tutorial program for each student based upon data about his aptitudes, abilities, and personality.

Teaching Program

A teaching program is made up of two parts: its logic and its content, the particular topic or set of topics to be taught. The complete set of rules used in making all of the decisions in teaching a student is the logic of his instruction. The content depends upon the student's pretest performance on a test covering the behavioral objectives of the instruction and his performance while learning. Selection from the library depends upon pretest performance. If pretest performance on a topic is already equal to the minimum acceptable final performance level, then the material used to teach it is eliminated from the student's program. Thus, the content of the student's program depends upon a contingency. His program contains only those materials which are necessary to teach the knowledge and skills which his initial level of performance indicated were inadequate in comparison with the terminal standards.

The functions involved in individualizing an instructional program are described in Stolurow and Davis (1965). Of those represented the more critical functions are the comparison of the student's responses with those expected (performance standards), the selection of particular displays, and the recording of a response history for on-line use as well as for course revision.

Teaching Strategy

A teaching strategy can be described in terms of two types of rules used in presenting a set of frames. The first type of rule determines sequence, or order, i.e., organization of the frames relating to an objective. The order in which instructional materials are presented to the student is determined by contingency rules relating to aptitude patterns. Each element of instruction is presented as a frame which requires a response (Markle, 1964). There is a variety of frames, each of which has a different purpose. The order in which they are arranged can make a difference in rate of learning and in retention (Stolurow and Lippert, 1964). Consequently, the type of frame and the sequence in which frames are arranged represent important specifications for a program. The second type of rule specifies the conditions under which the frames are presented; for example, a picture which is part

of a frame may be timed, or the medium used to encode the frame might be a film, for some, or an audio tape, for others. A complete set of rules includes both types, and the set employed in teaching a student to reach an objective is referred to as a teaching strategy. For example, a young student with high mathematics aptitude might be taught with a RULEG sequence under the conditions of the correction procedure and with visually displayed frames. An older student who is also high in mathematics aptitude might be taught the same concept with a RULEG sequence under the conditions of the non-correction procedure with audio tape. Each would be taught by a different teaching strategy.

Data for instructional decisions. In order for the teaching program to be effective, the system must use appropriate data correctly. The kinds of data to be used are specified by the teaching strategy, or rules expressed as contingency statements. To make individual predictions regarding the final performance of each student from his performance history, both the data and the rule possibilities have to be represented in the computer memory and used by the control program. In this respect the system operates in the tutorial mode just as it did in the pre-tutorial mode.

If the predictions about final performance are below the desired level, the system should be capable of changing the rules it is using and the basis upon which the rules are selected. For example, regarding the latter point, the system after teaching several students using a particular teaching logic, should use the data they provided it to re-analyze and compute new limits.

Whenever the student's performance does not meet the specified standards, adjustments are made in some part of his projected program. Also the data contribute to the adjustments made in the decision rules themselves for future applications to other students. In this way his data contribute to the pool of information used in making decisions about teaching subsequent students (e.g., Smallwood, 1962; Lippert, 1967). As a result of its previous teaching, the system learns how to teach more effectively.

Two Levels of Tutorial Instruction

Two levels of tutorial instruction are possible. The lower level obtains when the system is programmed to teach with a fixed logic and a fixed content, e.g., a linear program (Skinner, 1954), drill and practice (Atkinson, 1967). The higher level obtains when the system is programmed to alter either its logic, its content, or both, for subsequent parts of the instructional program depending upon the student's performance on the earlier parts. An example of the lower level is using the logic of a linear program (Skinner, 1954); another is using the logic of an intrinsic program (Crowder, 1960). For both of these strategies the content and logic must be specified in advance. At this level all of the instructional possibilities are pre-specified including the types of feedback given all students. At this lower level of tutorial instruction a CAI system provides a pre-specified learning environment which in one case is predetermined (linear program) and in the other case is contingent on the previous response.

A distinction can be made between a pre-specified learning environment and a learning environment that unfolds. In the former, the actual displays seen by the student are specified in kind and in sequence. If the student makes response "A," then he sees display α, etc. The system operates on the basis of conditionals that are of the form "If this response, then that stimulus display." All the conditionals are specified, and all students who make response "A" see display α. In the learning environment that unfolds, all students who make response "A" do not see display α, only some of them do. What display a student sees is determined by more than one rule, so that display α, for example, could be seen by some students who made response "A" but not by others who also made that response. An example of unfolding might be to show some students a frame containing the steps in the long division algorithm worked out with the values 360 divided by 12, while others may see the steps of this algorithm worked out with 480 divided by 52; some others may see examples with decimals. Still others may see the example worked out in a completely different sequence, e.g., successive subtraction and tallying. To do this on some basis other than chance implies that the decision is based upon more information about the student than that contained in the last response. This can be done by making available a response history of two, or more, previous responses. To do this a CAI system must be capable of keeping a cumulative record of responses in a form that makes the pertinent set available every time a contingency rule is to be used.

In addition to the availability of response histories, the CAI system programmed for unfolding a learning environment also must store in readily available form the rules for sampling the frames and the medium (e.g., audio or visual) or media to be used. Another requirement is the ability to use some contingencies, but not all, with a particular student. Related to this is the requirement for the system to be able to change the contingencies for each student if his performance is not meeting expected standards. This is called the feed forward concept in control theory. The system looks ahead to expected results based on what has happened and what possibilities could eventuate. After looking at the future possibilities it selects the most desirable one and then proceeds to do those things necessary to bring it about.

One type of rule specifies the sample of frames. In a linear program the sampling rules are simple: all students take the same program and they all see the same set of frames in the same order. The sample rule is to use the entire set of frames.

In idiomorphic programming the learning environment is predetermined but not pre-specified. A predetermined learning environment can be generated in either one of the two different ways referred to in the previous paragraph or by using both ways. Each way uses conditional rules of the form, "If response 'A,' then use procedure α to determine the next display." The display that appears next is predetermined by a procedure which can generate not just one, but a number of different displays.

One way to implement this form of conditional is to use a procedure that varies the content (e.g., the

numerals) but not the sequence (e.g., the algorithm). A second way to implement it is to use a procedure that varies the sequence, but not the content. A third way is to vary both. The first way makes the sequences or logic conditional; the third way makes both content and sequences conditional.

Another type of instructional strategy rule determines the time intervals. With a pre-specified learning environment of the linear type the strategy rule is simple since all students go from frame 1 to frame 2, to frame 3, etc., and in the same order. The rule is simple but the value can vary since the time intervals are determined by the student; he sets his own pace. By pre-specifying the conditions the only individualization achieved by a linear program is in the rate at which each student progresses through the material. By pre-specifying the outcome of each response in an intrinsic program, all the branching possibilities are fixed, and each response alternative takes a student to a specific display. The contingency that is pre-specified in this form of programming is that between the last response and the next display.

The learning environment can be predetermined in still other ways. For example, the content of the next frames may be different depending upon the correctness of the student's responses to other frames on the same topic, and their sequence may depend upon aptitude scores.

With the predetermined level of functioning, the on-line decisions made by a teaching system change one, or more, of the teaching rules, or the teaching strategy. Two students, each making the same response to frame 105, for example, might see different displays, experience different time intervals, and have information available for different lengths of time, etc. The information display, the sequence of displays in the set, and the timing conditions would be contingent upon specific information about the student's past performance. One question to be answered is what type of information can be used in making decisions. This is a research question to be answered ultimately by empirical data. Hypotheses can be formulated; for example, it is possible to include aptitude and personality test scores as well as sets of students' responses to instructional frames during learning in the "if" portion of the conditional statements. Whichever one, or combination of scores is used in specifying contingencies for instructional decisions will depend upon the educational programmer's hypothesis regarding the optimum strategy of instruction.

When a computer-based system provides a contingency-determined learning environment based upon individual differences in performance, it uses idiomorphic programming. In teaching a set of concepts the rules which determine the order in which they are presented may change, for example, from an inductive sequence to a deductive one. While any rule may be changed, it might be useful to consider examples in which only one is changed. For instance, only the rules determining the information sequence might be changed. Another example might be the rule specifying the evaluative feedback contingencies (given the correctness of the learner's responses). If the student's performance is not up to expectation in correctness or in response latency, the system may change the rules used to teach him. For example, the rule being used may have been to give the student reinforcement every time he responded within a certain period of time. If he did not respond quickly enough, the reinforcement may be given on only half of the times he responded. This type of teaching rule would specify the contingencies between response latency and the subsequent reinforcement schedule. If the level of the student's performance is below that expected, easier frames may be presented to him. This last example illustrates yet another type of rule, one making the difficulty of the subsequent material contingent upon the number of errors.

These examples illustrate the types of decisions that can be made when contingencies are predetermined but not pre-specified. Instruction is more individualized when instructional rules are contingent upon both the student's performance and what is expected of him. Whenever one or more aspects of the student's actual performance do not match expected performance, a new rule is used. When a new rule is used, sets of displays and/or frame sequences become available that would not be available if the system operated at the pre-specified level of a linear or an intrinsic program.

Bases for Contingencies in Instruction

This analysis of teaching as the control of the learning environment is based upon the idiographic contingency model of tutorial instruction (Stolurow, 1965b, 1965c). This model distinguishes two types of contingencies. At the lower level there are event contingencies; at the higher level there are rule contingencies. The model is made operational in SOCRATES in three ways.

Student data possibilities. The first is in the number of possibilities which the system provides for the use of data in making decisions concerning the content and strategy to be used in the instruction of individual students. For instance, the student's ability, aptitude, and personality, as well as his knowledge or skill, can be used in making decisions at any point in the program. Patterns of responses, that is, various patterns of error and time, made during learning are included in the possible student characteristics which the system can use. Sets of these patterns are possible bases for making the subsequent experiences contingent upon past performance. It is possible to use pretest performance along with error and time data on selected frames as the basis for content and/or sequential contingencies. Consider an example of how these data might be used to make different decisions for each of two students. One student may have made a high pretest score and then gone through the program rapidly, making lots of errors; another student may have made a low pretest score, but gone through the program in the same way. The first student might be assumed to be careless, and, therefore, given instruction designed to make him more careful. The second student could be assumed to have failed to learn because of insufficient ability and therefore, to be in need of supplementary or remedial instruction.

Different display possibilities. The earlier example in which the system provided different consequences to students who made the same response to a particular frame illustrates a second way a computer-based system can make instruction more individualized: namely, by providing different display possibilities for every pedagogically-significant contingency. The system's adaptivity is increased by including in its library a large number of possible displays for each response possibility. The number of possible displays is increased in different ways. One way is by writing sets of materials at different levels of difficulty, another is by using different forms of encoding to communicate information, e.g., verbal versus mathematical. A third way is to separate the informative material that is intended to explain the answer from the different kinds of evaluative feedback. The example of the student who was responding carelessly is a situation in which the availability of separate types of feedback would be useful. With the separate types available in this system, it would have been possible for the system to tell him he was careless while also telling the other student who made the same response that he was improving. Furthermore, each student could be given his subsequent instruction with another teaching strategy. For example, the careless one could be required to make correct responses before being allowed to go on. The other student in the example, however, could be branched to different evaluative feedback, possibly of an encouraging type, and then given easier material so he would have a higher probability of being correct.

Either an instructional theory or empirical data specifies each of the relationships between a set of student characteristics and a set of display possibilities; consequently, the more display possibilities in the memory of the system, the more contingencies it can handle and the more individualized the instruction could be.

The differences between the first and the second way in which adaptivity is achieved lie in the range of unique display possibilities which each contingency rule could generate. Furthermore, the actual number of display contingencies is larger than the number of frames since both the temporal interval between a response and a display and the duration of each display can be controlled separately. Additional contingencies are possible through the preparation of alternative materials at different levels of difficulty or of readability so that comprehension and speed can be maintained at each student's optimum level. For example, the careless student might be shown a display for a longer period of time before the system accepts his next response, or he may have to wait longer for the next frame after he has responded.

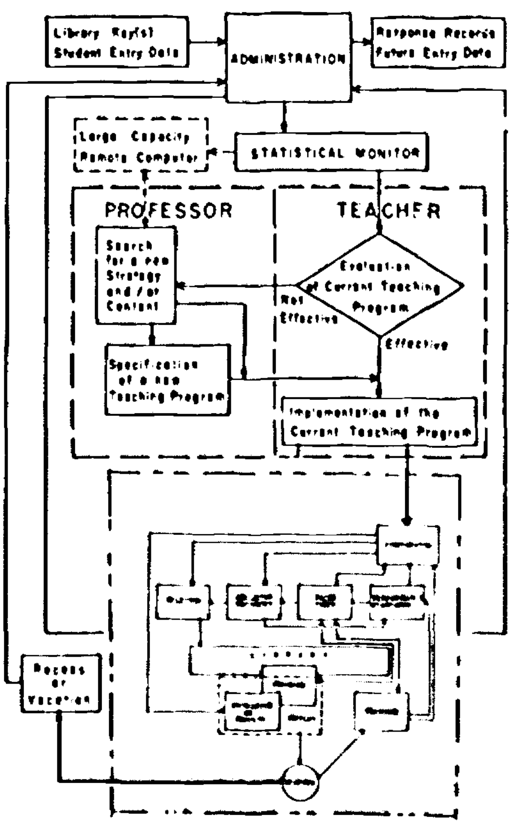
By programming the following conditions as contingencies, they can be used selectively in teaching: first, the difficulty or readability level of the information frames; second, their sequence; third, the temporal interval between responses and feedback as knowledge of results; fourth, the type of feedback statement used in terms of both their qualitative and evaluative nature; fifth, the schedule used to give feedback to the student; and sixth, the time interval between knowledge of results and the new program frame.

Criterion possibilities. A third way in which idiographic programming can make a system more adaptive is by using a variety of criteria in making decisions to either maintain or change the instructional rules themselves. A flow diagram describing a system for doing this is presented in Figure 3.

The program which allows the system to change rules on line is called the professor function. Basically this is the means for determining rule contingencies. It is to be contrasted with the ability of the system to function according to a fixed set of rules, which is called the teacher function. The teacher function is the part of the executive program that specifies the conditions for showing each one of a set of displays. The professor function is the part of the executive program that specifies the conditions for showing one set rather than another. These two functions are represented in Figure 3.

The professor function also requires "if... then..." rules, but these are rules that make the use of lower order rules contingent upon specific

Figure 3. System of criteria for decision-making in idiographic programming.



conditions. The professor rules imply teacher rules. In other words, if a student performs either better or worse than expected when a particular teaching strategy has been used, then some, or all, of these rules would be changed. This means that at least one other teaching strategy must be available as an alternative to the one being used. When an alternative is used, frames may be displayed that would not be displayed if the previous strategy had been used continuously.

An example of a pair of strategy rules which could be used by a system on an either/or basis would be correction procedure with an inductive sequence rule or with a deductive sequence rule. In using these strategies, it is necessary to have the conditions specified and a criterion available so that it is possible to determine if the student failed to meet the performance criterion. Assuming that he failed to meet the criterion when he was taught by the inductive strategy, he would be shifted to a deductive strategy in the next set of frames. In accomplishing this change, the criterion for the professor function might be ten consecutive correct responses none of which takes more than half a minute.

Whenever the student is being taught by one strategy rule, the teacher function is in control. However, the shift from one teaching rule to another is accomplished by the professor function which becomes effective whenever student responses fail to meet a criterion. Every time the student's performance fails to meet the professor criterion, one or more rules of instruction are changed. If a teacher criterion is not met, then the decision that is made is specified by a particular rule. The difference between the teacher and professor functions can be made more explicit by indicating that the use of one teaching rule may utilize a fixed set of instructional and feedback frames whereas the use of two or more rules may utilize additional informational and/or feedback frames. Therefore, with an executive program that can make rule changes and use different criteria for the changes, students may see materials when one rule is being used that they do not see if another one is being used.

Two types of criteria can be used in making decisions about instructional rules. One type of criterion is a predetermined value; it may be a value for errors, for time, or for a combination of both of these. It is illustrated by a conditional statement such as "If the student makes three or more correct responses to this set of frames, then continue with the use of rule x." The other type of criterion is generated by an algorithm, or computational formula, on line. This latter type of criterion is not predetermined; rather the algorithm and the nature of the data it uses are predetermined. Student response data are processed as they occur and the result of the processing is the criterion value used in deciding to continue or discontinue using a rule. The following is an example using response rate. If the student's rate of response increases so that the values fit a linear incremental function within a specified range of variation in slope, the strategy that is being used to teach him is maintained. If his rate of response does not fit a linear function within the range of permissible variation, the strategy shifts to the use of stronger positive evaluative feedback. Another example of a computed criterion is the use of a mean value based upon the performance of students who have just been taught. As each student is taught, his data are used in the re-computation of the mean and standard deviation which become the new criterion in making a decision with the next student. In this sense the system has a memory, for it accumulates data on students it teaches and it adjusts its future performance based upon the way in which it is programmed to use the accumulated data. These are some of the ways in which SOCRATES can utilize response history and criteria to determine its future operation. Thus it functions as a cybernetic system. A complete record of each response and its latency is available. These data can be summed across topics or across students as desired to obtain a performance measure that is read out or used internally by the system for educational or research reasons. Furthermore, all records are kept for a student rather than for a station. In other words, a student can move from one station to another and the system keeps all his records together. He can stop and start up again at any time and the system has immediate access to all of his test scores and past responses that it was programmed to save.

IMPLEMENTATION OF THE MODEL

In order to accomplish the teaching functions of the idiographic contingency model, it is necessary to prepare instructional materials, including tests, and to write computer programs to manage the instructional materials in order to collect the data and to process it.

LOADING THE SYSTEM

The preparation of instructional materials for use in SOCRATES involves two processes. One is photography; the other is computer programming. The student looks at photographed images of printed or drawn materials. These can be relatively conventional programmed instruction frames, since pre-tutorial decisions can be based either upon extant standardized tests of achievement, ability, aptitude, and personality, or upon specially-prepared tests. A convenience of SOCRATES, which does not exist for some other computer-based systems, is that any test material prepared in conventional form can be photographed and used on the system. Processing materials to make them actually available under computer control can be accomplished relatively quickly since they simply need to be photographed and then a program can be written to control the duration and sequence of their display. The students' responses are recorded and scored automatically.

When materials are loaded on SOCRATES there are two parts to the content of its library. One is the displayable information stored on 35 mm. film for which there is a maximum reel capacity of 1,500 frames; the other is the program in the memory of the computer, both core and disk, in which the instructional strategies and student data are stored. It is necessary to write a computer program to control the presentation of each frame of film and to process any data to be used on-line. With computer-based systems that use a CRT display, however, it

is necessary to write programs which generate the test items and the displays of program frames.

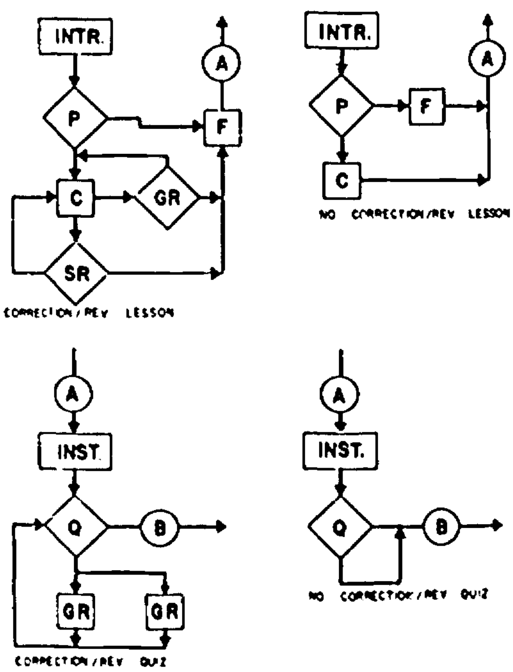
A more interesting use of SOCRATES occurs when educational programs are prepared especially for the system. It is possible to introduce a great deal more variation in the logic of instruction with the system than otherwise would be possible. Figure 4 presents a set of very simple flow diagrams associated with some research conducted with programs especially prepared for SOCRATES (Merrill and Stolurow, 1966). It also illustrates some of the discriminations that are made among materials to provide opportunities for flexibility in programming. P frames are those frames which represent programmed materials. F frames are feedback frames. C frames are frames which correct the student. GR frames are frames which provide for general review of the point covered in the P frame, and SR frames are frames which give specific reviews to the student. By providing these different types of frames in the film library, it is possible to route the student through selected subsets of them in whatever sequence seems most desirable. A computer program must be written to do this. It would be written to provide for the contingencies of a set of instructional rules.

The upper left panel of Figure 4 is a flow diagram in which a set of these frames is used to provide correction and review in a lesson. By correction procedure, we mean that the student must make a correct response to the material covered in the P frame before he is allowed to go on to the next P frame. If, however, it is desirable to route him through a review frame on this same material when the student responds in a way which the educational programmer feels indicates a review, the computer program is written to do this. The no-correction procedure is illustrated in the upper right panel. In this case the student is allowed to go on to the next frame whether he is right or wrong; however, he is first shown an F frame which tells him he is either correct or incorrect. The use of a no-correction rule is typical of a linear programming strategy.

The lower half of Figure 4 is a flow chart of the correction procedure and of the no-correction procedure as each is applied to questions in ordinary multiple-choice testing. The Q frame, or question, now replaces the P frame. The arrangement of P and Q frames on the film can be set up in whatever manner seems most efficient for use with the strategies that are intended for the students. The location of the frames in relation to each other on the film makes it possible to locate quickly the frames that could follow each P frame. The reason for this is to minimize the interval between the presentation of one frame and the retrieval of the next.

The decision rules to accomplish the various strategies are frequently presented initially in flow charts such as those in Figure 3. A group of one or more teaching rules may be required to implement a teaching strategy. Similarly, a set of decision rules or of teaching strategies comprises a teaching logic. The use of one or more strategies can be thought of as being employed at different points in the student's program to define the teaching strategy used with that program. If a particular set

Figure 4. Program flow diagrams illustrating the experimental conditions of correction/review and no correction/review on lessons and quizzes.

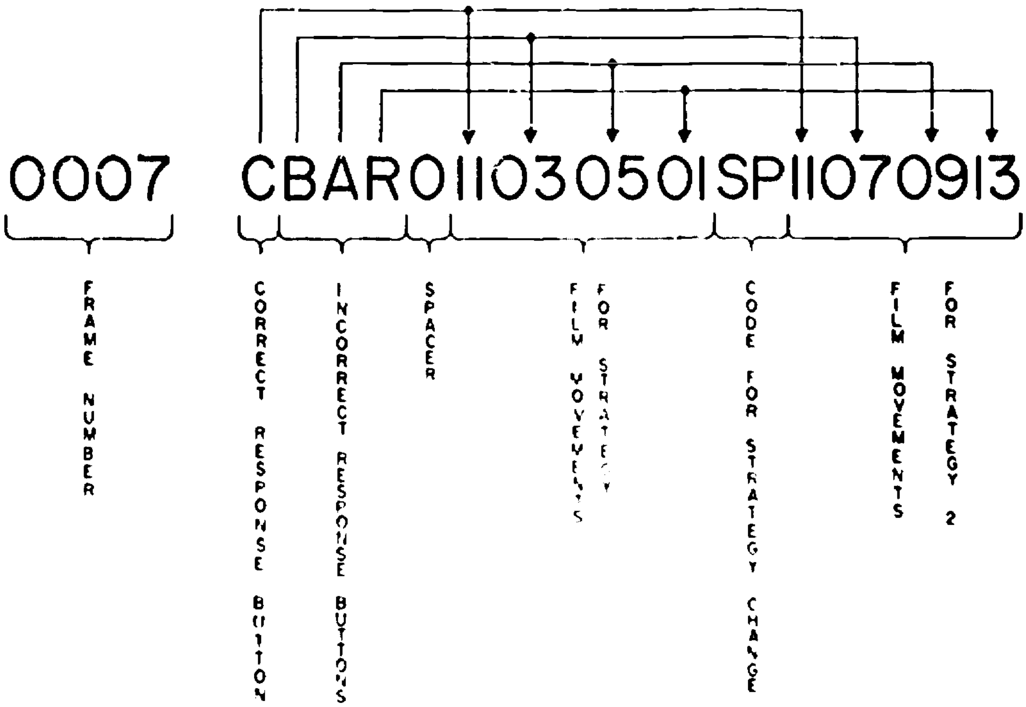


of contingency rules is used to provide the instructional strategy and a set of one or more strategies is used to make up the teaching logic, then the system is more or less adaptive to individual student needs depending upon the extent to which the particular set used to teach a student was determined before the student started. If the set was determined prior to instruction, the system is not as adaptive as if it were determined while he was learning and depended upon his performance during learning.

The research problem in making the system adaptive in an effective way is to determine the contingencies to use and the criteria so that the computer-based instructional system is programmed to implement those contingencies that instruct most efficiently. This is the primary problem in the further development of the idiographic contingency model. In order to solve this problem it will be necessary to begin by using the systems in a less adaptive level than they are capable of performing. Students with different characteristics will be assigned to strategies to determine their relative performance.

Code line of library table. A code line of a library table is presented in Figure 5. It represents one of the two sets of computer materials used to control instruction. The library table consists of lines of code which determine the frame locations and the various legal buttons which the student may

Figure 5. A code line from a library table.



use in responding to displayed material. The illustrated code line presented in Figure 5 is for instructional frame number 7 of a program. This is indicated on the left by the number 0007. Associated with each instructional frame is a set of lettered alternatives each of which is specified by a rule that is part of a teaching strategy. Each letter in the code line is associated with a frame address to which the student will be taken if he makes that response. If, for example, he makes the correct response, "C" in this example, the film is moved eleven framed forward. "C" is a correct response. The correct response alternatives are always listed before the incorrect ones, and after the first one the other correct alternatives are starred. There is only one correct response in the example. After "C," the letters "B" "A" "R" appear. Each is an allowable incorrect response. If any other key is used it results in no change in the display.

In addition to the library table there is the computer program which actually manages the movement of the program frames and the times of display. The various strategies and levels of adaptivity previously described can be accomplished through the use of these two sets of materials.

RESEARCH

Some of the research which relates to the idiographic model follows. While the model specifies the types of contingencies that appear to be sufficient for the design of adaptive learning environments, the exact nature of the contingencies which will maximize the rate of learning remains to be determined by research. Some studies which have provided information about contingency relationships can be described briefly. Since they are insufficient to document all of its facets, the studies mainly provide support for the general validity of the model.

THE EFFECTS OF SEQUENCE UPON APTITUDE REQUIREMENTS FOR LEARNING

A matrix of inter-correlations between a particular set of aptitude scores and the gain scores resulting from a learning experience indicates the extent to which the gains can be predicted from the various aptitude scores (DuBoise, 1962). The significance level of each correlation indicates whether the aptitude is or is not a significant factor in determining the gain score which students make. If students are randomly assigned to different teaching strategies, and a single set of aptitudes is related to the gains made, then differences in the correlations indicate that an aptitude is involved in learning under one strategy but not under another. This type of information relates to the diagnostic aspect of the model in that the correlations would identify the aptitudes that could be used in assigning students to the instructional strategy that will result in the greatest gains. If the strategies consist of rules that determine the frame sequence, then the students'

aptitudes could be used to determine their optimum sequence.

One characteristic of teaching materials which can be made contingent upon aptitude is the sequence of frames presented to the student. If studies show that the size of the correlations between particular aptitudes and learning scores varies when different sequences of P frames are used so that some students do well when they learn with one sequence while others do well with a different sequence, it can be assumed that there is a rule that determines the sequence which would be best to use for a student who is relatively high in a particular aptitude and relatively low in another. Research revealing that aptitude-gain correlations do change when different strategies are used provides support for the idiographic contingency model. If aptitude-gain correlations do not change significantly, then this type of model would be inappropriate.

Asynchrony as a measure of sequence. One problem that arises in using a diagnostic-prescriptive approach is the lack of critical information. Critical information might be obtained through the use of variables in research that can be implemented in a selective way. For example, if frame sequence is to be manipulated selectively then it is important to have the rules for manipulating sequences so formulated that they can be implemented easily. Therefore, the way in which the sequence of frames, or displays, is characterized is important. The concept of asynchrony as applied to critical and noncritical stimulus elements appears to be quite useful for this purpose (Detambel and Stolurow, 1956). Asynchrony can be exemplified in the following illustration involving color, size, and shape. If all frames in a program use color, size, and shape of a set of objects and the student is learning a concept involving one of them, such as color, then color is the critical cue. Consider some of the things that happen in going from frame to frame in a sequence. The color may change from one frame to the next—for example, from red to blue—but the size and shape may not. In this example, not only are color and size asynchronous with respect to one another, but also color and shape are asynchronous. If this pattern of asynchrony existed in a program, then as the learner went from one frame to the next, color would change but neither size nor shape would change. It is highly unlikely that this set of relationships would continue indefinitely, but that is not the point being made at this time. The point being made is that for a series of frames it is possible to look at changes in critical characteristics and determine what other characteristics also change or do not change. In this way it is possible to determine the number of pairs of changes in stimulus characteristics which take place in going from one frame to the next one. If one characteristic, such as color, is critical to what is being taught, and other characteristics are not, then two different types of asynchronous sequences can be identified. One consists of asynchronous sequences in which a critical feature changes but a noncritical one does not. This can be called a CC type of asynchronous sequence since the critical characteristic changes. A second asynchronous sequence exists when a critical feature does not change but a noncritical feature does. This can be called a CNC type of asynchronous sequence.

Two opposite types of sequences also can exist. These are called synchronous sequences. One type occurs when both a critical feature and a noncritical feature change; the other occurs when neither changes, or, in other words, the display is repeated. These ways of characterizing sequences have been found to be useful in organizing materials in several studies (e.g., Detambel and Stolurow, 1956; Stolurow, 1956; Stolurow, Fogel, and Fogel, 1966; Anderson, 1966) and recently in Germany by Kotter (1965).

These concepts are treated more extensively in Stolurow (1956) and in Garner (1962).

The use of class-descriptive cues. The measures of asynchrony, while reliable for a particular frame sequence, do not, however, identify a unique sequence of frames. Using the same set of frames it is possible to make up different sequences which have the same amount of asynchrony. In other words if a sequencing rule were used by a computer to generate an asynchronous sequence the computer would not necessarily generate the same sequence each time the rule was used. However, in all cases the frequency with which CC and CNC types of sequences occurred would be the same. Since this is so, it is useful to differentiate among sequences with the same asynchrony value. A way of doing this is to use the concept of class-descriptive cue (Stolurow, 1956); any feature that does not change from one frame to the next frame within a subset of two or more frames of a program can function as a class-descriptive cue. Therefore wherever there is more than one feature that can be useful to the learner, the frames should be organized into subsets with a class-descriptive cue that is related to one of the objectives of instruction. The data thus far collected indicate that the use of class-descriptive cues in organizing displays does facilitate learning (Stolurow, 1956; Wulff and Stolurow, 1957).

The effects of sequence changes on the relationship between aptitudes and gains. An asynchrony rule in combination with a cue-organizing rule can produce sequences that provide class-descriptive cues for the learner. This teaching logic appears to be a useful way to synthesize a program of instruction.

Two studies will be reported here to indicate the results of this approach. In both studies analyses were made of the data to determine the resulting relationships between aptitude and gain scores when different sequences of the same frames were used for instruction.

The first study was conducted by Cartwright (1962). He used a fractions program which was presented to mentally-retarded students during secondary-level instruction. Two variations of the instructional material were prepared. In one program there was a systematic increase and decrease of the numerator and then the denominator of the fractions presented to the student as he went from frame to frame. In the other alternative, which was made up with the same frames, the sequence used did not maintain this relationship. The comparison of the two groups was therefore between a program with a

TABLE 1

CHANGES IN CORRELATIONS BETWEEN APTITUDE AND LEARNING SCORES FOR DIFFERENT SEQUENCES WITH EQUAL AMOUNT OF ASYNCHRONY (CARTWRIGHT, 1962)

Individual Characteristics	Sequence with Systematic Increment and Decrement (N = 20)	Sequence with Non-systematic Increment and Decrement (N = 16)
Mental Age - WISC		
Verbal	.02	-.26
Performance	.26	-.50*
Total	.17	-.49*
Mental Age - Binet	.18	-.49*
Language Aptitude	.12	-.54*

*$p < .05$

class-descriptive cue and one without a class-descriptive cue. Some of the data are reported in Table 1.

Cartwright found, for example, using a total intelligence test score from the WISC, that a .17 correlation existed with final achievement or gain scores under the class-descriptive cue (CDC) condition; whereas a correlation of -.49 was achieved with a sequence that did not have a class-descriptive cue. With the Binet intelligence test scores comparable correlations were obtained, namely, .18 and -.49. Even more interesting is the pair of correlations with the verbal aptitude test: a positive correlation of .12 was obtained with the CDC sequence and a -.54 with the other sequence. These data indicate that verbal aptitude becomes more or less significant as a contributor to learning depending upon the sequence in which the frames of the program are presented.

The difference in the correlations for the two sequences indicates that it is possible within a 30-point IQ range, from 53 to 83 in this case, to predict from either the general intelligence measure or the verbal aptitude measure what the residual scores of students will be under the non-CDC sequence. This means that with the CDC sequence neither intelligence nor language aptitude plays a significant role in determining the residual scores within this range of ability. With another sequence, however, these abilities played a more important part in determining what the students learned.

A second study, dealing with sequence differences and also employing a fractions program, but a different one, was conducted with normal students in elementary schools in Champaign, Illinois. This study was completed recently and revealed a pattern of relationships between aptitudes and learning as a result of sequence differences which was similar to that reported in the Cartwright study. Table 2 summarizes the set of correlations between pairs of variables which differed significantly for the two degrees of asynchrony. While there were other significant differences between the effects of these two sequences in addition to those shown in this table, these differences illustrate the point that changes in asynchrony can alter the aptitude requirements for achievement.

The data from both studies support the idiographic contingency model described earlier. They indicate that effective individualization of instruction can be accomplished by using teaching rules that offer instructional sequences which fit the aptitude characteristics of each student. The use of sequencing, and particularly of asychronous sequencing with CDCs needs to be studied further to determine the sequencing rules which are so related to the students' aptitudes that the best match for maximizing his performance is achieved. The research problem facing us now is to determine the specific kinds of relationships one needs to take into account in making the best match between aptitudes and rules for generating particular sequences of frames.

THE EFFECTS OF VARIATIONS IN ENCODING ON APTITUDE REQUIREMENTS

Another aspect of a teaching strategy is the form of symbol used to express a set of concepts or, in other words, the encoding employed to communicate the concepts to students. A recent study (Stolurow, Frase, and Odell, 1966) of encoding differences employed two different sets of notation in teaching logic to college students. One program used the Peano-Russell notation system, the other the Polish notation system. In all other respects the two programs were identical. Thus, any variations in outcome could be attributed to the encoding differences.

The first set of results (Table 3) indicates rather clearly that under the Peano-Russel notation system American students perform in a manner that is predictable from a set of aptitude tests. Comparable

TABLE 2

CORRELATIONS BETWEEN IQ AND PERFORMANCE AND BETWEEN SPECIFIC APTITUDES AND PERFORMANCE FOR DIFFERENT DEGREES OF ASYNCHRONY (UNPUBLISHED STUDY)

Variables	Sequence with Less Asynchrony	Sequence with Higher Asynchrony
California Test of Mental Maturity		
Total IQ and Gain Score	+.01	+.73*
Non-verbal IQ and Gain Score	-.13	+.55*
Iowa Test of Basic Skills		
Spelling and Gain Score	-.32	+.56**
Stanford Achievement Test		
Arithmetic and Gain Score	-.46	+.70**

* and **: The two correlations are significantly different from each other; *$p < .05$; **$p < .01$

TABLE 3

CORRELATION OF DIFFERENT APTITUDES WITH PERFORMANCE ON A TEST COVERING IDENTICAL CONTENT TAKEN AFTER LEARNING THE POLISH OR PEANO-RUSSELL NOTATION

Group	Vocabulary	Aptitudes for Mathematics	Inductive Reasoning	Inference
Peano-Russell	.27*	.31**	.31**	.01
Polish	.09	.18	-.01	.01

*$p < .05$ **$p < .01$

students, when trained with the Polish notation system, however, achieved gains which did not correlate significantly with the same aptitude tests. Thus, it appears that these aptitude measures are predicting the processing skills which are required of the learner by the symbol system used to encode the concepts and not the actual learning of the logical concepts themselves. This seems to be a rather clear finding since the concepts of logic were identical in the two programs. The processing skills which are required to decode the symbols representing these concepts appear to be quite different. Decoding the Polish expressions requires that a person trained to process English and mathematics statements be able to transpose symbols in order to make sense out of the displays presented to the students. These data indicate that one can teach the same set of concepts to students by two different, but presumably equivalent, encoding systems and in doing so can change not only the aptitude requirements for learning but also the transfer effects which the training produces. The data also indicate that the Peano-Russell notation system is superior to the Polish system for students trained in processing the English language.

THE EFFECTS OF VARIATIONS IN EVALUATIVE FEEDBACK ON PERSONALITY REQUIREMENTS

Still another set of contingencies pertaining to the model are those relating personality characteristics of the learner to his performance gains when different conditions of evaluative feedback, or social reinforcement, are employed. Social reinforcement in these studies means evaluative feedback statements which tell the student about the quality as well as the correctness of his response. Thus, approval could be expressed in evaluative statements such as, "Excellent, you are coming along fine," or "Right on the button"; for negative evaluation, or reproof, evaluative statements such as "No, you are not thinking," "No, what a silly mistake to make" express the attitudes.

Each of these kinds of statements was given to different groups of students all of whom studied the

TABLE 4

CORRELATIONS BETWEEN PERSONALITY VARIABLES (a) PERFORMANCE, (b) ATTITUDE AND (c) TIME UNDER FOUR DIFFERENT CONDITIONS OF SOCIAL REINFORCEMENT (FRASE, 1965)

Group	Aggression/ Performance	Aggression/ Attitude	Deference/ Attitude	Deference/ Time
Condition 1 ±	.15	.24	-.05	.07
Condition 2 +	-.44*	-.50*	-.06	.48*
Condition 3 -	.35	.73***	-.59*	.37
Condition 4 0	-.50*	-.16	.00	-.40

$^{*}p < .05$ $^{***}p < .005$

same basic program in logic (Frase, 1963). Group one (±) had correct and incorrect responses evaluated. Group two (+) had only their correct responses evaluated. Group three (-) had only their incorrect responses evaluated. Group four (0) had no evaluation of either one, just knowledge of results.

In Table 4, three sets of correlations are of interest. One set is between personality test scores and learning scores (e.g., aggression/performance); a second is between personality scores and attitude (e.g., aggression/attitude); and a third set is between personality and learning time. All three sets revealed that each of the four experimental conditions of social reinforcement was most suited to a different personality group. Three personality characteristics, aggression, deference, and intraception (Murray, 1938) were found to be important in relation to performance under the four different conditions of social reinforcement. For example, students who were high in aggression performed best when they were given negative evaluation of their wrong responses and not given positive evaluation of their right responses. Under two conditions – number 2, positive evaluation, and number 4, no social reinforcement – aggression was related negatively to performance. In other words, for the students high in aggression, performance was impaired under these conditions. Under condition number 3, negative reinforcement, aggression was related positively to attitude but unrelated to achievement. But the relationship with performance, although positive, was not significantly different from zero.

Under the punitive conditions (number 3) high deference was related to lower attitude but was not related to achievement performance.

Intraception was negatively correlated with speed and efficiency, and related to high attitude under maximum social reinforcement, condition number 1.

In another study by Parisi (1965) the basic paradigm of the Frase study was replicated and an additional personality test, the Cattell 16 PF, was given prior to learning. All materials were administered in Italian to Italian students in Italy. These results support those of Frase that negative evaluations were more important that positive evaluations. Furthermore, positive reinforcement was more effective than negative reinforcement in reducing the correlation between intelligence and performance. Performance on the achievement test was related to deference, or lack of need for autonomy, when social reinforcement was used. When no social reinforcement was used, performance was related to need for achievement and exhibitionism.

The data from these studies seem to show that it is possible to alter the quality and speed of the performance as well as the attitude of students by the way in which social reinforcement is used during the learning experience. They further suggest that it might be possible to achieve an optimum matching of personality characteristics with social reinforcement conditions for each individual student so that the general level of effectiveness of instruction can be raised by individualizing these contingencies.

SUMMARY

This paper summarizes a general model used to design a computer-based instructional system. This idiographic contingency model defines classes of variables that are presumed to be important in adaptive instruction. It also specifies relationships between those variables that characterize the learner before and after learning which can be used selectively to individualize instruction. Included among the "before" measures are knowledge, aptitude, and personality. The data indicate that the "before" measures have different relationships with the "after" measures (e.g., amount learned and attitude) under different conditions of learning. These data support the model.

The kinds of decisions that need to be made to individualize instruction were discussed and a set of research studies relating to them was described briefly. Relevance of the data to the model was indicated and it was pointed out that learning environments can be made more adaptive by using appropriate matching of: (a) aptitude and sequence contingencies; (b) aptitude and encoding contingencies;

TABLE 5

CORRELATIONS BETWEEN INTELLIGENCE SCORES AND PERFORMANCE VARIABLES - (a) MIDPROGRAM TEST, (b) POSTTEST AND (c) ERRORS IN PROGRAM - UNDER FOUR DIFFERENT CONDITIONS OF SOCIAL REINFORCEMENT (PARISI, 1965)

	16 P. F. Factor B[b] Intelligence with			Verbal Reasoning with		
Group[a]	LM	LP	E	LM	LP	E
Condition 1 ±	-.07	.33	-.22	-.15	.05	-.24
Condition 2 +	-.16	.51*	-.17	.40	.37	-.31
Condition 3 -	.38	.49*	-.44	.40	.49**	-.23
Condition 4 o	.48*	.42*	-.46*	.47*	.46*	-.18

[a]The social reinforcement conditions are coded as follows: Condition 1± mean positive and negative social reinforcement; Condition 2+ is positive only; Condition 3- is negative only; and Condition 4 o is neither positive nor negative.

[b]LM is logic midprogram; LP is logic posttest; and E is errors.

*$p < .05$ **$p < .01$

and (c) personality and evaluative feedback contingencies.

The technological capability for accomplishing a high degree of individualization in instruction exists. The problem now is to find the critical psychological and educational data which would make such a technological capability an effective means of instruction.

FOOTNOTE

1. This paper appeared as Technical Report No. 12 from the Training Research Laboratory, Office of Naval Research, Contract NONR 3985 (04). Urbana, Illinois: University of Illinois, 1966.

REFERENCES

Anderson, R. C. Two rules for manipulating the sequence of stimuli in concept learning tasks. Journal of Educational Psychology, 1966.

Atkinson, R. C. Instruction in initial reading under computer control: the Stanford project. Journal of Educational Data Processing, 1967, 4 (4), 175-192.

Bushnell, D. D. Computer-based teaching machines. Journal of Educational Research, 1962, 55, 528-531.

Cartwright, G. P. Two types of programmed instruction for mentally retarded adolescents. Unpublished master's thesis, University of Illinois, 1965.

Coulson, J. E. (Editor) Programmed learning and computer-based instruction. New York: Wiley, 1962.

Crowder, N. A. Automatic tutoring by intrinsic programming. In A. A. Lumsdaine and R. Glaser (Editors), Teaching machines and programmed learning: a source book. Washington, D. C.: National Education Association, 1960. Pp. 286-298.

Detambel, M. H. and Stolurow, L. M. Stimulus sequence and concept learning. Journal of Experimental Psychology, 1956, 51, 34-40.

Duboise, P. and Glaser, R. (Editors), Training research and education. Pittsburgh: University of Pittsburgh Press, 1962.

Frase, L. T. The effect of social reinforcers in a programmed learning task. Technical Report No. 11, Training Research Laboratory, University of Illinois, 1963, Contract NONR 1834 (36), Office of Naval Research.

Garner, W. R. Uncertainty and structure as psychological concepts. New York: Wiley, 1962.

Gentile, J. R. The first generation of computer-assisted instructional systems: an evaluation review. University Park, Pennsylvania: Computer Assisted Instruction Laboratory, Pennsylvania State University, 1965.

Hansen, D. N. Computer assistance with the educational process. Review of Educational Research, 1966, 36, 588-603.

Kötter, L. Der Einfluss unterschiedlicher Signalfolgen aud den Konzepterwerb (minimale und maximale Variation der Signale). Bericht über den 24. Kongress der Deutschen Gesellschaft fur Psychologie. Göttingen: Verlag fur Psychologie, Dr. C. J. Hogrefe, 1965.

Lippert, H. T. Computer aided learning and transfer effects of Russian pronunciation. Unpublished doctoral dissertation, University of Illinois, 1967.

Markle, S. M. Good frames and bad: a grammar of frame writing. New York: Wiley, 1964.

Merrill, M. D., and Stolurow, L. M. Hierarchical preview versus problem oriented review in learning an imaginary science. American Educational Research Journal, 1966, 3, 251-261.

Murray, H. A. Explorations in personality. New York: Oxford, 1938.

Parisi, D. Social reinforcement and performance in programmed learning in Italy. Technical Report No. 27, Training Research Laboratory, University of Illinois, Contract NONR 1834 (36), Office of Naval Research, 1965.

Silberman, H., and Coulson, J. E. Automated teaching. In H. Borks (Editor), Applications in the behavioral sciences. Englewood Cliffs, New Jersey: Prentice Hall, 1962. Pp. 308-335.

Skinner, B. F. The science of learning and the art of teaching. Harvard Educational Review, 1954, 24, 86-97.

Smallwood, R. D. A decision structure for teaching machines. Cambridge, Massachusetts: M. I. T. Press, 1962.

Stolurow, L. M. Utilization of class-descriptive cues in the learning of technical information - studies in task engineering. In G. Finch and F. Cameron (Editors), Symposium on Air Force human engineering, personnel and training research. Technical Report 56-8. Baltimore: Air Research and Development Command, 1956. Pp. 248-266.

Stolurow, L. M. Teaching by machine. Cooperative Research Monograph No. 6, OE-34010. Washington, D. C.: U.S. Office of Education, 1961.

Stolurow, L. M. A model and cybernetic system for research on the teaching-learning process. Programmed Learning, 1965, 2, 138-156. (a)

Stolurow, L. M. Idiographic programming. National Society for Programmed Instruction Journal, 1965, 4, 10, 12. (b)

Stolurow, L. M. Untersuchungen zur kybernetischen Modellierung von Lehr und Lernprogrammen. Zietschrift für Psychologie, 1965, 171 (suppl.), 247-274. (c).

Stolurow, L. M. The computer-assisted instructional system in theory and research. In D. Unwin and J. Leedham (Editors), Aspects of educational technology. London: Methuen, 1966. Pp. 257-273.

Stolurow, L. M. Computer-based instruction: psychological aspects and systems conception of instruction. Journal of Educational Data Processing, 1967, 4(4), 193-215.

Stolurow, L. M., and Davis, D. Teaching machines and computer-based systems. In R. Glaser (Editor), Teaching machines and programmed learning, II: Data and directions. Washington, D. C.: National Education Association, 1965. Pp. 162-212.

Stolurow, L. M., Fogle, R. E., and Fogle, C. A. A comparison of two degrees of asynchrony. Technical Report No. 10, 1966, Training Research Laboratory, University of Illinois, NDEA Title VII Contract 4-20-002, U.S. Office of Education.

Stolurow, L. M., Frase, L. T., and Odell, S. J. Relative transfer effects of learning systems of notation for encoding expressions in a course of programmed symbolic logic. Unpublished manuscript, Training Research Laboratory, University of Illinois, 1966.

Stolurow, L. M., and Lippert, H. T. Prompting, confirmation and overlearning in the automated teaching of a sight vocabulary. In J. P. De-Cecco (Editor), Educational technology. New York: Holt, Rinehart and Winston, 1964. Pp. 187-197.

Taber, J. I., Glaser, R., and Schaefer, H. H. Learning and programmed instruction. Reading, Massachusetts: Addison-Wesley, 1965. Pp. 65-72.

Wulff, J. J., and Stolurow, L. M. The role of class-descriptive cues in paired-associates learning. Journal of Experimental Psychology, 1957, 53, 199-206.

Zinn, R. Computer assistance for instruction. Automated Education Letter, 1965, 1, 4-14.

PART IV

Instructional Methods and Teacher Behaviors

TEACHING METHODS, or the system of interactions which occur between a teacher and students, have been the subject of many investigations in the last decades. In this Part, two recent paradigms for research on teaching are presented. N. L. Gage points out that the global research on teaching methods has not produced consistent, replicable or useable results. Therefore, a different and more limited approach must be used. In "An Analytical Approach to Research on Instructional Methods" he discusses micro-teaching, a technique which permits limited subsets of a comprehensive teaching system to be explored analytically. Micro-teaching permits the gathering of data on a limited set of student-teacher interactions during a short span of time, less than a class period. From such research "micro criteria" of effectiveness may be hypothesized and tested in terms of technical skills observable within the micro-teaching situation.

Another analytical approach to research on teaching is proposed in the paradigm set forth by Ned A. Flanders in "Interaction Analysis and Inservice Training." Flanders proposes a system of research in which a limited number of categories of student-teacher interaction is used in observing classroom interactions. Interaction data are recorded at 3 second intervals. The resultant matrix of data allows rather effective analysis of the verbal interchange between the teacher and the pupils. Interaction analysis is discussed in terms of its use as a research tool and as a means of effecting behavioral changes in teachers, such as inservice programs.

An Analytical Approach to Research on Instructional Methods

Nathaniel L. Gage
Stanford University
Palo Alto, California

INSTRUCTIONAL method constitutes one of the most important and promising but also the most frustrating of the areas of educational research and development. Compared with the areas of learning, subject matter, instructional materials, and organization for instruction, instructional method appeals to the classroom teacher as closest to the heart of her problem. It is all right for a teacher to know about learning, to know his subject matter, to have appropriate instructional materials, and to fit into a given organization for instruction. But what a teacher really wants to know is, "What should I do in the classroom?" If you ask prospective teachers or teachers on the job, "Where do you really want help?" I think the reply will deal with *some* aspect of instructional methods.

Unfortunately, the urgency of this demand has not been accompanied by corresponding success in meeting the demand. Research on teaching has yielded relatively few solid and usable results. The stature of theory and research in other areas puts them well ahead of the study of teaching in the struggle for scientific maturity.

Implicit in what I am saying is a basic distinction between research on learning and research on teaching. The former deals with all the conditions under which learning, or a change in behavior due to experience, takes place. And as I have already indicated, the study of learning is relatively mature, well established, with many volumes of substantial literature to its credit. Research on teaching, on the other hand, deals with a subset of the conditions under which learning occurs in one person, namely, the conditions established by the behaviors of another person, called the teacher. As our schools have developed during the past two or three millennia, we have always attempted to promote and improve the learning process through the intermediation of such another person. Until very recently, the assumption that teachers were helpful or even necessary for many important kinds of learning that society wanted to promote went unchallenged. Even today, the challenge of independent study or computer-assisted instruction and other devices is a mere whisper against the thunder of the assumption that teachers are necessary, that teachers are here to stay. My discussion of instructional methods is going to be based on that assumption. It is the relatively neglected, undernourished, and underachieving subset of research on learning which I call research on teaching that I shall deal with here.

RESEARCH ON TEACHING

The Stanford Center for Research and Development in Teaching is devoted to this problem area. In its conceptual framework, teaching, or the behaviors and characteristics of teachers, stands at the center. This domain contains variables that serve as both independent and dependent variables in the Center's program. When the teachers' behaviors and characteristics serve as independent variables, and the pupils' learning and behaviors serve as the dependent variables, then we have research on teacher effectiveness, or, more neutrally, research on teacher effects. On the other hand, when teachers' behaviors and characteristics serve as the dependent variables, and teacher education programs and procedures serve as the independent variables, then we have research on teacher education. Taking

both research on teacher education and research on teacher effects as our domain, we have research on teaching, as it is understood in the Stanford Center. My subject is the Center's program of research and development on instructional methods, and my procedure will be to work from the past to the present, from the general to the specific, and then to try for a look at where we are going. When I get down to specifics, I shall be dealing primarily with research in which I am involved, rather than all of the research underway in our Center.

PAST RESEARCH

But first, let us look at where research on teaching has been. As the behavioral sciences go, it has a respectably long history but a regrettably inglorious one. Research on teaching has been going on almost as long as research on learning. Some studies were made in the 1910's and 1920's, and quite a few were made during the 1930's. By the early 1950's, substantial reviews and bibliographies of research on teaching began to appear. And during the last decade, the flow of research on teaching has indeed become significant. But the early years did not pay off in solid, replicable, meaningful results that had considerable theoretical or practical value. Positive and significant results were seldom forthcoming, and they survived replication even less often. The research yielded many findings that did not make sense, that did not hang together in any meaningful way.

Under these conditions, as Kuhn (1962) has pointed out, research workers are impelled to reexamine their first principles, the paradigms by which they guide their efforts. The model problems and solutions of the community of researchers on teaching were accordingly subjected to more and more reappraisal. Licking the wounds inflicted by their negative results, researchers on teaching built up a modest literature of new conceptual frameworks, approaches, and paradigms for research on teaching.

To illustrate, let me refer to one of the dominant paradigms that even today leads many discussions and research projects into the wilderness. This is the paradigm that says that what we need above all, before we can select and train better teachers, is the criterion of teacher effectiveness. Here is one example of that kind of approach:

> The lack of an adequate, concrete, objective, universal criterion for teaching ability is thus the primary source of trouble for all who would measure teaching. One typical method of attack used in rating scales is to compile a list of broad general traits supposedly desirable for teachers, with respect to which the rater passes judgement on each teacher. This amounts to an arbitrary definition of good teaching, which is subjective and usually vague, but it does not necessarily lead to an identification of it. Only if the traits themselves can be reliably identified can their possessor be identified as a "good teacher" according to the definition laid down in the scale. Even when the scale is made quite specific, relating not to general traits but to concrete procedure, the fundamental difficulty remains, that there is no external and generally accepted criterion against which the scale can be validated to establish the significance of its items (Walker, 1935, pp. x-xi).

This kind of writing implies that there is some magic variable that applies to all of teaching, for all pupils, at all grade levels, in all subject matters, and in all objectives. The phrase "the criterion of teacher effectiveness" betokens a degree of generality that has seldom been found in any branch of the behavioral sciences. It also reflects the mistaken notion that such a criterion, largely a matter of values, can be established on the basis of scientific method alone.

RECENT RESEARCH

The so-called criterion problem misled a whole generation of researchers on teaching, embroiled them in endless and fruitless controversy, and lured them into hopelessly ambitious attempts to predict teacher effectiveness over vast arrays and spans of outcomes, teacher behaviors, time intervals, and pupil characteristics, all on the basis of predictive variables that had only the most tenuous theoretical justification in the first place. It is little wonder that, when Berelson and Steiner (1964) dealt with the subject of teachers' behaviors and characteristics in their inventory of scientific findings in the behavioral sciences, they dismissed the "large number of studies" with the single dismal sentence that "there are no clear conclusions" (p. 4[illegible]).

If the global criterion approach had proved to be sterile, what was the alternative? The answer was to take the same path that more mature sciences had already followed: If variables at one level of phenomena do not exhibit lawfulness, break them down. Chemistry, physics, and biology had, in a sense, made progress through making finer and finer analyses of the phenomena and events they dealt with. Perhaps research on teaching would reach firm ground if it followed the same route.

Apparently, a number of students of the problem had this general idea at about the same time. In 1962, writing my chapter on paradigms for the Handbook of Research on Teaching, I coined the term, "micro-criteria" of effectiveness. As I said in that chapter:

> ... One solution within the "criterion-of-effectiveness" approach may be the development of the notion of "micro-effectiveness." Rather than seek criteria for the over-all effectiveness of teachers in the many, varied facets of their roles, we may have better success with criteria of effectiveness in small, specifically defined aspects of the role... a sufficient number of laws applying to relatively pure aspects of the teacher's role, if such laws could be developed, might eventually be combined, ... to account for the actual behavior and effectiveness of teachers with pupils under genuine classroom conditions" (Gage, 1963, p. 120).

A group of workers at Stanford University, to which I was to move a few months later, took a similar view. In the Stanford program for training secondary-school teachers, Robert Bush, Dwight Allen, and their co-workers adopted what is now known as the technical skills approach. Technical skills are specific instructional techniques and procedures that a teacher may use in the classroom. They represent an analysis of the teaching process into relatively discrete components that can be used in different combinations in the continuous flow of the teacher's performance. The specific set of technical skills adopted in the teacher-education program at Stanford may be quite arguable. Indeed, the list of skills has been revised a number of times over the past few years. What is important is the approach – the attempt to analyze teaching into limited, well-defined, components that can be taught, practiced, evaluated, predicted, controlled, and understood in a way that has proven to be altogether impossible for teaching viewed in the larger chunks that occur over a period of an hour, a day, a week, or a year.

When analyzed-teaching, in the form of technical skills, is made the focus of our concern, we find it possible to do fairly satisfying research both on teacher education and on teacher effects. The satisfaction comes from being able to measure or manipulate relevant independent variables, perform true experiments, or make careful analyses and measure relevant dependent variables.

The idea of technical skills may be illustrated by the terms used in a recent list of such skills. One was called "establishing set," or the establishment of cognitive rapport between pupils and teacher to obtain immediate involvement in the lesson; one technique for inducing a positive set is the use of relevant analogies. A second technical skill is that called "establishing appropriate frames of reference," or points of view. A third technical skill is that of "achieving closure," or pulling together major points, linking old and new knowledge, at appropriate points within a teaching episode as well as at the end. A fourth technical skill is that of "using questions" in such a way as to elicit the kinds of thought-processes and behaviors desired such as simple recall, or concept formation, or evaluation. Other technical skills are those in "recognizing and obtaining attending behavior," "control of participation," "providing feedback," "employing rewards and punishments," and "setting a model."

MICRO-TEACHING

These technical skills into which important aspects of the teaching job have been analyzed are not merely the subjects of lectures and discussions in the teacher education program. Rather, they form the basis for the intern's practice teaching prior to his entrance into actual classrooms. This procedure, well known by now as "micro-teaching," consists in getting the trainee to teach a scaled-down teaching exercise. It is scaled down in terms of time because it lasts only 5 to 10 minutes. It is scaled down in terms of class size, because he teaches a group of not more than five pupils, who are brought in and paid to serve as pupils in the micro-teaching clinic. It is scaled down in terms of the task, since the trainee attempts to perform only one of the technical skills in any single micro teaching session. The sessions are recorded on video tape, and the trainee gets to see and hear himself immediately after the session. While he looks at and listens to himself, he receives criticisms and suggestions from supervisors trained to be both perceptive and tactful. Then he "re-teaches" the same lesson to a new small group of pupils in an attempt to improve on his first performance of the specific technical skill that is his concern in that session.

Obviously, the general idea is subject to many variations. The size of the class can be manipulated; the number of trainees teaching a given group of children can be increased; the duration of the lessons can be lengthened; and the nature of the teaching task can be made more complex so as to embrace a group of technical skills in their real-life combinations. But the idea of analyzing teaching into technical skills remains the heart of the method and provides its power as a paradigm for research.

The research on micro-teaching and technical skills in the Stanford teacher education program has taken the form of experiments in which various procedures for feedback to the trainee are manipulated. Professors Dwight Allen and Frederick McDonald have organized a program of research on variables hypothesized to influence the learning of the technical skills of teaching. Their independent variables fall into three categories: practice variables, feedback variables, and demonstration variables. A practice variable may consist in micro-teaching versus teaching in an actual classroom. A feedback variable may be the positive or negative character of the feedback, or the mediation of the feedback by another person rather than the trainee himself. Finally, a demonstration variable may take the form of symbolic demonstration, consisting of written or spoken words, or perceptual demonstration, consisting in either live or video-taped portrayals of the desired behavior; and each of these can consist either of self-modeling or modeling by others. Other independent variables have been identified, such as the timing of a reinforcement, the amount of practice, and the amount of feedback.

This condensed description of the Allen-McDonald research program can suffice to illustrate the use of the analytic approach to research on teacher education. Their research takes the form of true experiments in which subjects are randomly assigned to different values of the independent variable.

TECHNICAL SKILLS APPROACH – EXPLAINING ABILITY

I should like to turn now to an example of the way in which the technical skills approach can be applied to the study of teacher effects. This research has dealt with a technical skill that I call "explaining," or the skill of engendering comprehension – usually orally, verbally, and extemporaneously – of some process, concept, or generalization. Explaining occurs in all grade levels and subject matters, whether it is a fifth-grade teacher explaining why the time in New York differs from that in San Francisco or a geologist explaining how the ice age may have been

caused by volcanic eruptions. Everyday observation tells us that some people explain aptly, getting to the heart of the matter with just the right terminology, examples, and organization of ideas. Other explainers, on the contrary, get us and themselves all mixed up, use terms beyond our level of comprehension, draw inept analogies, and even employ concepts and principles that cannot be understood without an understanding of the very thing that is being explained. To some of us, it has seemed that explaining comes very close to being the inner essence of instruction, so that when a teacher is attempting to explain proportionality to his geometry class or irony to his English class, he is behaving more purely as a teacher than when he is attempting, say, to motivate, promote discussion, or maintain discipline. At any rate, we have made some studies of explaining ability in the attempt to determine some of the characteristics of effective explanations.

EXPLAINING ABILITY STUDY IN MICRO-TEACHING CLINIC

The first study was made in the micro-teaching clinic at Stanford during the summer of 1965 by J. C. Fortune, R. F. Shutes, and N. L. Gage (1966). We attempted to determine the generality of explaining ability, that is, the degree to which the ability to explain one topic was correlated with the ability to explain another topic, and the degree to which the ability to explain a topic to one group of pupils on one occasion was correlated with the ability to explain the same topic to another group of pupils on another occasion. We also were able to design the study so as to determine the degree to which there was generality over both pupils and topics, or the degree to which the ability to explain one topic to one group of pupils on one day correlated with the ability to explain another topic to another group of pupils on another day. Because there were only sixty pupils to be shared in groups of five among approximately forty interns in the micro-teaching clinic, the design became quite complex in order to avoid having any intern teach the same topic to the same group of pupils more than once and to avoid having the same group of pupils receive an explanation of the same topic more than once. Accordingly, the forty social studies interns – and we chose to work with the social studies interns only because there were more of them than any other kind of intern – were divided into five clusters of eight interns each. The lectures dealt with twenty different topics, each consisting of an "Atlantic Report" from the Atlantic Monthly. The correlations that we obtained were thus medians of five correlations, each based on four, six, six, six, and eight interns, respectively.

The index of lecture effectiveness, or what I would like to call the micro-criterion of teacher explaining ability, was the pupils' mean score on a 10-item test of their comprehension of the main ideas of the lecture, which was presented by each intern in 15 minutes under somewhat standardized conditions. This mean score was adjusted for the mean ability of the pupils in the given group as measured by their scores on all of the other topics. Similarly, any given mean score was adjusted for the difficulty of the topic as measured by the mean score of all groups of pupils on that topic. Thus, the variance of the adjusted mean posttest comprehension scores was attributable not to the ability of the pupils or the difficulty of the topic but rather to the differences among the teachers. We then investigated the question of the various kinds of generality by determining the median intercorrelations among the various means. The upshot of this part of the study was that generality over topics was nonexistent, and generality over groups was about .4. In other words, the interns were moderately consistent in their ability to explain the same topic to different groups on different occasions, but they were not consistent in their ability to explain different topics.

The study dealt with the correlations between explaining effectiveness and the pupils' rating of various aspects of the explanations. The pupils rated the interns' performance with respect to twelve items, such as clarity of aims, organization of the lesson, selection of material, and clarity of presentation. To us it seemed that some of these dimensions should correlate more highly with explaining ability than others. In particular, we hoped that such discriminant validity would be manifest in the form of a higher correlation between the mean rating of the lecture for "clarity of presentation" than for any of the other items of the Stanford Teacher Competence Appraisal Guide. Our hope was borne out; the correlation of the adjusted mean posttest comprehension scores with pupils' ratings of "clarity of presentation" was .56, higher than that with any of the other rating scale items. This result seems to us to support the validity of both the index of lecture effectiveness and the mean ratings by the pupils.

EXPLAINING ABILITY STUDY IN PUBLIC SCHOOLS

During the school year 1965-1966, I was able, in collaboration with Barak Rosenshine, to undertake a replication and extension of this study in the public schools (Rosenshine, 1963). Because there was no lack of students in the high-school classes, taught by their own teachers, we did not become involved in the complexities of design necessary in the micro-teaching clinic. To put it very briefly, we got forty eleventh-grade social studies teachers each to deliver a 15-minute lecture on an "Atlantic Report" on Yugoslavia taken from the Atlantic Monthly. The teachers had been given the article several days in advance, and had been told to prepare a lecture that would enable their pupils to answer a 10-item multiple-choice test of comprehension of the article's main ideas. To guide them in preparing their lecture, they were given five of the multiple-choice questions that would be asked, while the other five questions were withheld. After the 15-minute lecture, in which the teachers were permitted to use the blackboard but no other aids, their students took the 10-item test. They also rated the teacher's lecture on items similar to those I have already described. The next day, the same teachers and classes did the same things, except that the subject matter was an "Atlantic Report" on Thailand; again the teachers had been given five of the ten items as a guide to the kind of lecture that they should prepare and had been told to focus on the explanation of the major ideas, concepts, and principles brought out in the article, which constituted the curriculum for this bit of teaching. On the third day, the classes heard

a third lecture, one that was the same for all classes, a tape recorded 15-minute lecture on Israel, a verbatim reading of an "Atlantic Report," and then the pupils again took a 10-item test based on that article.

The class mean on the Israel test was used to adjust the class means on Yugoslavia and Thailand for between-class differences in ability. Our reasoning was that the score on such a test of comprehension of a uniform lecture would be more useful in controlling relevant kinds of ability than would the usual scholastic aptitude test. The class means on Yugoslavia and Thailand were also adjusted for teacher differences in the content-relevance of the lecture, as determined by scoring the transcript of the lecture for relevance to the ten items on the comprehension test. We then assumed that the variance that still remained in the adjusted comprehension test means of the classes would reflect differences between the teachers in what we were concerned with, namely, the intellectual style and process of the teacher's lecture. In this study, the teacher's adjusted effectiveness index on Yugoslavia correlated .47 with his effectiveness on Thailand; i. e., there was considerable generality of effectiveness over topics, even after student ability and content relevance had been partialed out.

It should be noted that we were using the micro-teaching idea in this investigation. The teaching was restricted to just one aspect of the teacher's role, namely, ability to explain the current, social, political, and economic situation in another country. The curriculum was also scaled down. We also used another major feature of the micro-teaching clinic, the video tape recorders which made it possible for us to study the teacher's behavior, verbal and nonverbal, at leisure.

One major question was that of whether our criterion, or micro-criterion, of teacher effectiveness in explaining, namely, the mean comprehension score of the pupils, adjusted for both mean pupil ability and content-relevance, contained variance that would be manifested in something about the lecture that was visible or audible. In other words, was there some difference between good and poor explanations that was worth trying to analyze? So we picked two lectures on Yugoslavia that were extremely high on our index of effectiveness and two that were extremely low. We had a group of eight judges read the article on Yugoslavia and take the comprehension test, and then watch and listen to all four of these lectures. Then the judges ranked the lectures in terms of perceived effectiveness in engendering comprehension as measured by the 10-item test. It turned out that the judges' post-dictions were quite significantly more accurate than could have been expected on the basis of chance, and we were accordingly reassured that our micro-criterion was indeed reflected in something that could be seen or heard in the lecture.

But the major concern of this investigation was to determine the cognitive and stylistic correlates of the lecture's effectiveness. For this purpose, we used extreme groups to minimize the labor of scoring a host of variables about which we had no great conviction. So the ten most effective explanations on Yugoslavia were identified and also the ten least effective. From these, we chose at random five of the most effective and five of the least effective. Then, groups of judges and content analysts worked over the transcripts of the lectures, scoring and rating them on a host of variables. Some of these were sentence fragments, the average sentence length, the number of prepositional phrases per sentence, and so on. Other variables dealt with the number of self-references by the teacher, or with various aspects of syntax, or instructional set, familiarization, uses of previous knowledge, mobilizing sets, attention focusing procedures, organization, emphasis, amount of repetition and redundancy, the number of words per minute, and so on.

The variables that discriminated between the five best and the five worst lectures on Yugoslavia were then tried out on the other set of five best and five worst lectures on Yugoslavia to see if they still discriminated. Those that survived this first cross-validation were then tried out on the best and worst lectures on Thailand. At the last accounting, two characteristics of the lectures had survived this kind of validation and cross-validation procedure. These variables were what we are calling "explaining," or the degree to which the teacher describes the how, why, or effect of something, and the "rule-eg-rule" pattern, or the degree to which the teacher states a generalization, gives examples of it, and then summarizes a series of illustrations at a higher level of generality than the illustrations themselves. These two variables not only seem to be valid in our data but also are reliably rated by independent judges. Nonetheless, these must not be considered to be firmly established findings; they are merely examples of the kinds of conclusions to which research of this kind can lead.

Currently, we are in the process of scoring all of the explanations on all of the variables that appear to hold any promise, and we will then undertake studies of the complete correlation matrices involving not only the indices of explanation effectiveness, but also all of the characteristics of the explanations, and the ratings of the lectures by the pupils who heard them. Such a complete correlational study will throw light on the consistency from one lecture to another of the indices of lecture effectiveness, and the stylistic characteristics of the lectures, and also their intercorrelations.

What I have been describing is of course a correlational study. Along with its advantages in permitting the exploration of a wide variety of possible correlates of explaining ability as they occur under fairly normal conditions, it also has the disadvantage of making causal interpretations hazardous. For this reason, studies of this kind ought to proceed fairly rapidly into experiments in which the different ways of explaining will be based, at least in part, on leads obtained from our correlational studies.

Such experimental research may lead toward quite novel methods of teaching that could never be developed on the basis of studies of teaching the way teaching is. Stolurow (1965) has contrasted the approach of "modeling the master teacher" with that of "mastering the teaching model." The first approach is that of studying the most effective teachers

we can find in order to find out how they behave and what they are like so that we can attempt to produce more teachers like them. Many research workers see little promise in this approach. They recommend that we undertake instead to develop wholly new models of the teaching process designed for optimal effectiveness regardless of their similarity to the way teaching now goes on in the normal classroom.

COMPUTER-ASSISTED PROGRAMMED INSTRUCTION

The teaching model that many advocates of this approach have in mind is that of programmed instruction, particularly computer-assisted programmed instruction. As Suppes and Atkinson and others have described this revolutionary undertaking in research and development on instruction, it holds out great promise indeed. Before too long, the annoying problems in the hardware will have been solved. After a somewhat longer time, we may expect substantial and well-validated programmed curricular materials to have been developed in all the subject matters and grade levels. As one who has seen the highly developed installations at the Brentwood School in East Palo Alto, California, I must share the optimism of Suppes and Atkinson, and other developers of computer-assisted instruction.

Their very success, or coming success, raises problems for the kind of instructional methods with which we have been concerned in this paper. On superficial examination at least, certain major problems of ordinary classroom teaching seem to be clearly surmounted by computer-assisted instruction. For example, the problem of the cognitive complexity in the teacher's task, of how the teacher can say just the right thing at the right time to develop a concept or formulate a theory, is apparently well handled, at least in principle, by computer-assisted instruction. Its programs can be worked out and tried out in meticulous detail, well in advance, at leisure, by the most skilled curriculum experts in the land, and then made available in all their subtlety and complexity to every teacher who uses the program. Another major problem in the ordinary classroom is that of individualizing instruction. No matter how we group our pupils between schools, within schools, or within classrooms, we still have the problem of adjusting the rate and direction of the teaching and learning process to the needs and abilities of the individual pupil; here again, at least in principle, computer-assisted instruction seems at first glance to have the better of the live teacher.

While pondering these problems, I got some help from a restatement of the idea of individualized instruction in a recent paper by Philip Jackson (1966). He states it as follows:

> Individualizing instruction, in the educator's sense, means injecting humor into a lesson when a student seems to need it, and quickly becoming serious when he is ready to settle down to work; it means thinking of examples that are uniquely relevant to the student's previous experience and offering them at just the right time; it means feeling concerned over whether or not a student is progressing, and communicating that concern in a way that will be helpful; it means offering appropriate praise, not just because positive reinforcers strengthen response tendencies, but because the student's performance is deserving of human admiration; it means, in short, responding _as_ an individual _to_ an individual.

Individualization in this sense is much more than allowing for differences in speed of moving through a program or providing different branches or routes through the material.

Jackson's analysis of this kind of limitation in computer-assisted instruction should be placed alongside of the indications by Suppes (1966) that tutoring and dialog, which are higher levels of instruction than drill-practice, are still well in the future, as capabilities of computer-based instruction. Hence, any fears about the rapid obsolescence of live teachers, even where narrowly defined cognitive objectives are concerned, are quite unwarranted. That is, there will still be a need for teachers to use the kinds of technical skills, including explaining, with which the analytic approach being developed at Stanford and elsewhere is concerned. We shall have to continue to grapple with the problems of cognitive complexity and individualization through the medium of the live, human, teacher even in the realm of the well-formulated cognitive objectives. And there will always be the indispensable role of teachers in assisting pupils in attaining various kinds of affective and social learnings in the classroom.

Accordingly, group discussions, role playing, teaching for divergent thinking, as well as the technical skills I have already mentioned, are all the subject of various research and development projects now under way in the Stanford Center for Research and Development in Teaching. We are also looking at the way in which the organizational context influences the teacher's choices among ways to teach. And, in one of our projects, entitled "The Teacher in 1980," we are looking at the way in which new curriculum developments, television and other technical aids, computer technology, and new organizational schemes in the schools will affect the teacher's role in the foreseeable future.

CONCLUSION: A NEED FOR ANALYSIS

In conclusion, let me refer again to what I see as one basic new theme in the research and development in teaching that is now under way at Stanford and elsewhere. If it were necessary to sum it up in one word, my word would be <u>analysis</u>, breaking down the complexities that have proven to be so unmanageable when dealt with as a whole. We are no longer crippled by the notion that because there is one word called "teaching," there is one, single, over-all criterion of effectiveness in teaching that will take essentially the same form wherever teaching occurs. Even if none of the analyses of teaching that we have now proves to be viable, they will not be replaced by the old global conceptually impossible, complex variables that I see as the reason for the fruitlessness of so much of research on teaching in the past.

Instead, they will be replaced by other analyses of teaching, perhaps even finer analyses, until we get the sets of lawful relationships between variables that will mark the emergence of a scientific basis for the practice of teaching. It may well be that a 15-minute explanation of a 5-page magazine article is still too large a unit of teaching behavior to yield valid, lawful knowledge. It may well be that the mean score on a 10-item test of comprehension, adjusted for student ability and content relevance of the lecture, is still too large and complex a dependent variable. But, compared with the massive, tangled, and unanalyzable units that have typically been studied in the past – in research on the lecture method, the discussion method, and class size, for example – such units seem precise and manageable indeed. And eventually, of course, we shall have to put teaching back together again into syntheses that are better than the teaching that goes on now. I think it would be safe to say that there is now some hope of our being able to develop a scientifically grounded set of answers to every teacher's central question, "What should I do in the classroom?"

REFERENCES

Berelson, B. and Steiner, G. A. Human behavior: An inventory of scientific findings. New York: Harcourt, Brace, and World, 1964.

Fortune, J. C., Gage, N. L., and Shutes, R. E. Generality of the ability to explain. Paper presented at the meetings of the American Educational Research Association, Chicago. February, 1966.

Gage, N. L. Paradigms for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963. Pp. 94-141.

Jackson, P. W. The teacher and the machine: Observations on the impact of educational technology. Unpublished manuscript, University of Chicago, 1966.

Kuhn, T. S. The structure of scientific revolutions. Chicago: University of Chicago Press, 1962.

Rosenshine, B. Behavioral correlates of effectiveness in explaining. Unpublished doctoral dissertation, Stanford University, 1968.

Stolurow, L. M. Model the master teacher or master the teaching model. In J. D. Krumboltz (Ed.), Learning and the educational process. Chicago: Rand McNally, 1965. Pp. 223-247.

Suppes, P. The uses of computers in education. Scientific American, 1966, 215 (3), 207-220.

Walker, Helen M. Preface. In W. H. Lancelot, A. S. Barr, T. L. Torgerson, C. E. Johnson, V. E. Lyon, A. C. Walvoord, and G. L. Betts, The measurement of teaching efficiency. New York: Macmillan, 1935. Pp. ix-xiv.

Interaction Analysis and Inservice Training

Ned A. Flanders
The University of Michigan
Ann Arbor, Michigan

WHAT IS INTERACTION ANALYSIS?

INTERACTION analysis is a system for observing and coding the verbal interchange between a teacher and his pupils. The assumption is made that teaching behavior and pupil responses are expressed primarily through the spoken word as a series of verbal events which occur one after another. These events are identified, coded so as to preserve sequence, and tabulated systematically in order to represent a sample of the spontaneous teacher influence.

The most important criterion which any coding system must meet before it can be considered satisfactory is that a trained person can decode the data in order to reconstruct those aspects of the original behavior which were encoded, even though he was not present at the observation. A part of this article will describe inferences which can be made from a blind analysis of coded data.

Interaction analysis has been used to study spontaneous teaching behavior and it has also been used in projects which attempt to help teachers modify their behavior. In the first instance there may be a long period of time between observations and the analysis of the data. The data can be punched on IBM cards as they are collected over a period of several months, but a computer program to tabulate and analyze the data may be used only after all the observations have been completed. On the other hand, when interaction analysis data are collected in order to provide a teacher with information about his own behavior as a part of preservice or inservice training, then it may be advantageous to code directly into a desired tabulation form so that interpretations can be made at the earliest moment after the teaching episode is completed. This article will be more concerned with the procedures of interaction analysis which can be used during preservice and inservice training, and less concerned with applications in more basic research projects.

A possible goal of an inservice training program might be to discover whether the spontaneous patterns of verbal communication which are observed are, or are not, consistent with the intentions of the teacher. In such a program, the assumption is made that modifying behavior in order to make it more consistent with intent will, in most instances, result in an improvement. The model toward which behavior is modified is created by the teacher attempting to change.

An inservice training program can also be designed in which all participating teachers attempt to make similar modifications, for example, developing more skill in making full use of the ideas which are expressed by pupils during classroom discourse. Here the model could be justified from the results of research which made use of interaction analysis, but special care would be necessary in designing activities which would help teachers accept and understand the desirability of such modifications. In brief, the value orientation would be that the quality of classroom instruction is improved when the ideas expressed by pupils are more adequately recognized, clarified, used in some step of a logical

analysis, thus giving the pupils a more active part in the learning activities.

In nearly all applications of interaction analysis it is desirable to collect other kinds of data such as pupil attitudes, pupil achievement, and perceptions of the teaching situation held by the teacher and the pupils. These additional data permit the development of theory and explanation.

The resources of such a program, as with all inservice activities, require money in order to obtain time, space, and assistance for those who are participating. A number of steps can be anticipated: (a) Prospective observers and participating teachers must be trained to use the technique. This usually requires 6 to 12 hours under the direction of a qualified observer. (b) Personal, self-development goals must be clarified by the participating teachers. (c) Each teacher-observer team must set up comparison situations involving a planned change of behavior and two or more observations should be scheduled so that evidence of change can be assessed. (d) Plans to collect other types of information must be completed. (e) Social-skill training sessions, based on interaction analysis categories, are often helpful, especially when they are incorporated in the observation training. Such training often helps teachers set personal goals for changing their behavior. (f) Plans will be necessary to provide incentives for teachers and to maintain the momentum of the program, once it is started. (g) Care must be taken to insure freedom from threat, to make sure that the voluntary aspects of participation are genuine, and to avoid superimposing the program as an excessive demand on teachers who may already be too busy.

OBSERVATION PROCEDURES AND MATRIX INTERPRETATION

Given ten categories, shown in Table 1, all verbal statements are classified at least once every 3 seconds by a trained observer. The events are coded by using the arabic numbers from one to ten which are written down in such a way as to preserve the original sequence. The data can then be tabulated in a table of ten rows and ten columns which is called a matrix.

Such a series is entered into a matrix two at a time. The first number of each pair indicates the row of the matrix, the second the column. The first pair consists of the first two numbers. The second pair consists of the second and third numbers, and thus overlaps the first pair. All tallies enter the matrix as a series of overlapping pairs.

With one tally approximately every 3 seconds, there are one hundred tallies for 5 minutes, 1,200 tallies per hour; therefore, 20 minutes, or about four hundred tallies, provide a matrix with sufficient data for a number of inferences about verbal communication.

In a sustained observation of a teacher covering six to eight 1-hour visits, it is necessary to tabulate separate matrices for different types of classroom activities. Each matrix should represent either a single episode of class activity or any number of homogeneous episodes that are combined. We use five activity categories for junior high school academic subjects; they are: routine procedures, discussion of new material, discussion to evaluate student performance or products of learning, general discussion, and the supervision of seatwork or group activities. Different activity categories may be useful for a self-contained elementary-school classroom. In any case, the purpose of tabulating the data in several different matrices, instead of in just one total matrix, is determined by the purposes of observation and the range of expected classroom activities.

In the case of the inservice training of teachers, data are tabulated into separate matrices in such a way as not to mask the comparison to be made. To illustrate, suppose the comparison is between two samples of teaching behavior, one before and one after several weeks of social skill training. Keeping the data in homogeneous activity matrices will help to avoid false conclusions. For example, a decrease, increase, or no change in certain categories, when the two matrices are compared, may be due to the differences in the learning activities observed, rather than due to change in teaching behavior which resulted from inservice training. Grouping the data to represent homogeneous learning activities helps to avoid such confusion.

A tabulated matrix divides into special areas for interpretations that are shown in Table 2. Particular questions can be answered by comparing tallies within and between these areas. Here are some examples.

Areas A (1+2+3+4), B(5+6+7), C(8+9), and D(10) can be used to find the percent time the teacher talks, the pupils talk, and time spent in pauses, silence, and confusion. Comparisons between Areas A and B provide information about the relative balance between initiating and responding within teacher talk. Initiating teacher talk is more directive, tends to support the use of teacher authority, and restricts pupil participation. Responsive teacher talk is more indirect, tends to share authority, and expands pupil participation.

Area E is a block of nine cells that indicates the continued use of acceptance and praise, constructive reaction to pupil feeling, and clarifying, accepting, and developing pupil ideas, as well as transitions among these three categories while the teacher is talking. In any inservice training program devoted to increasing the teacher's attention to ideas expressed by pupils, before and after comparisons would require an analysis of these nine cells. In fact, an inservice training program which attempted to teach more subtle differences in the teacher's reaction to pupil ideas might require subdividing Category 3 in order to note the presence and absence of various types of Category 3 statements. For example, 3-1 -- merely repeats to show that the pupil ideas were heard; 3-2 -- reacts to specific pupil ideas, but only in terms of the teacher's perceptions of these ideas; 3-3 -- reacts to specific pupil ideas, but reactions incorporate the perceptions of one or more pupils; and 3-4 -- stimulate a reaction to a pupil's ideas by asking questions so

TABLE 1

CATEGORIES OF INTERACTION ANALYSIS

Teacher Talk and Student Talk

TEACHER TALK

Response

1.* Accepts feeling: accepts and clarifies the feeling tone of the students in a non-threatening manner. Feelings may be positive or negative. Predicting or recalling feelings are included.
2.* Praises or encourages: praises or encourages student action or behavior. Jokes that release tension, but at the expense of another individual; nodding head, or saying "um hm?" or "go on" are included.
3.* Accepts or uses ideas of students: clarifying, building, or developing ideas suggested by a student. As teacher brings more of his own ideas into play, shift to category five.
4.* Asks questions: asking a question about content or procedure with the intent that a student answer.

Initiation

5.* Lecturing: giving facts or opinions about content or procedures; expressing his own ideas, asking rhetorical questions.
6.* Giving directions: directions, commands, or orders to which a student is expected to comply.
7.* Criticizing or justifying authority: statements intended to change student behavior from non-acceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-reference.

STUDENT TALK

Response

8.* Student talk – response: talk by students in response to teacher. Teacher initiates the contact or solicits student statement.

Initiation

9.* Student talk – initiation: talk by students which they initiate. If "calling on" student is only to indicate who may talk next, observer must decide whether student wanted to talk. If he did, use this category.
10.* Silence or confusion: pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

*There is NO scale implied by these numbers. Each number is classificatory; it designates a particular kind of communication event. To write these numbers down during observation is to enumerate, not to judge a position on a scale.

that other pupils react. In effect, Category 3 is expanded into four categories for a special purpose. This would result in a 13 x 13 matrix instead of a 10 x 10 matrix.

Area F is a block of four cells that indicates the continued use of directions and criticism and transitions between these two categories. The two transition cells are particularly reliable indicators of discipline problems. Shifting from directions to criticism is tallied in the 6-7 cell, and indicates that expected compliance is judged unsatisfactory by the teacher. Shifting from criticism back to directions, the 7-6 cell, indicates a return to more directions after criticism.

Areas G_1 and G_2 are particularly interesting because they isolate the immediate response of the teacher at the moment students stop talking. One aspect of teacher flexibility can be discovered by comparing the balance of indirect and direct statements shown in G_1 and G_2 with those found in Areas A and B. The difference between superficial, short, perfunctory praise or clarification, and praise or clarification that is more carefully developed is easily seen by comparing the tallies in Area G_1 with those in E, particularly the 2-2 and 3-3 cells.

Area H indicates the types of teacher statements that trigger student participation. Responses to the teacher are found in column 8; statements initiated by the student in column 9. As one might expect, there is usually a heavy loading of tallies in the 4-8 cell. High frequencies in this cell and the 8-4 cell, but not in the 8-8 cell, often indicate rapid drill.

Area I indicates sustained student participation. These may be lengthy statements by a few students, or student-to student communication.

So-called "steady state" cells fall on the diagonal from cell 1-1 to 10-10. Tallies here indicate

TABLE 2

AREAS OF MATRIX ANALYSIS

CATEGORY	CLASSIFICATION	CATEGORY	1	2	3	4	5	6	7	8	9	10	TOTAL
ACCEPTS FEELING	RESPONSE	1											
PRAISE		2		Area E									
STUDENT IDEA		3											
ASKS QUESTIONS	INITIATION	4								Area H			
LECTURES		5					"Content Cross"						
GIVES DIRECTIONS		6											
CRITICISM		7						Area F					
STUDENT RESPONSE		8		Area G₁				Area G₂		Area I			
STUDENT INITIATION		9											
SILENCE		10											
		Total	Area A				Area B			Area C		Area D	
			Response			?(s)	Initiation			Student Talk		Silence	

that the speaker persists in a particular communication category for longer than 3 seconds. All other cells are transition cells moving from one category to another.

Outlined in the center of Table 2 by dash lines is the content cross. The total number of tallies in this area, compared with tallies not in this area, gives a very crude indication of the content orientation of the class activity.

In addition to making use of the areas just described, the following procedure can be followed to interpret a matrix.

First, locate the single cell within the ten rows and columns which has the highest frequency. The pair of events, represented by the cell, is the most frequently occurring and can be used as a starting point in reconstructing the interaction.

Second, from this highest frequency cell, you start forward or backward, in terms of sequence, to begin a sequence diagram. The row of any cell indicates the most likely third event, that is, the event which is most likely to follow, given an original pair of events designed by the highest frequency cell. The column, on the other hand, indicates which event most probably preceeded the pair of events in question. The flow of events is properly represented when the eye scans the matrix in a clockwise rotation. Should the highest frequency fall into a transition cell, not a steady state cell, the row or column of either number in the pair can be studied to retrace or advance the sequence of events.

An example of matrix interpretation will be shown later in this article. Skill in matrix interpretation, however, is not likely to develop from reading this article, which serves only to propose guidelines. For that matter, skill in observation cannot develop from reading about how it is done. All aspects of interaction analysis require practice in order to develop skill. It is the opposite of a spectator sport.

FEEDBACK AND CONSULTATION WITH TEACHERS

The purpose of feedback is to provide a teacher with information about his verbal statements which permits a comparison with some standard or model of what should have happened. Without purposive

comparisons which are planned in advance of the observation, the reactions of the person receiving the information may be reduced to incidental speculations or points of interest that happen to be noticed.

To be useful as a model or standard, one's intentions must be specified in terms of frequencies to be found in the cells of the matrix. Thus, if a teacher wishes to practice providing more extended praise, he will expect to find an increase in the 2-2 cell of the matrix, one situation compared to another. If a teacher decides he would like to stimulate more pupil talk in which ideas are initiated by pupils, he might study column 9 in the matrix to see what events trigger these pupil statements. In a program of inservice training in which all teachers attempt to increase the utilization of ideas expressed by pupils, attention will be directed to column 3 and row 3 of the matrix.

Most inservice training programs can achieve some success in bringing selected concepts and value orientations to the attention of teachers, these are matters of awareness. Evidence that such awareness has been implemented through overt behavior requires an objective assessment of spontaneous teaching behavior.

The assessment of the spontaneous behavior must be reasonably objective in order to be reliable. Unfortunately, interaction analysis is not free of bias and error, probably about one out of every ten classifications of an experienced observer is incorrect. Interaction analysis data can be and probably are more objective, when dealt with in summary form, compared with most other procedures for making judgments about spontaneous teaching behavior. Judgments about events which occur within time segments of only a few seconds and which must be repeated again and again tend to become more consistent with practice. Furthermore, noting the presence or absence of a short event is not a procedure which lends itself as easily to distortion and bias.

No matter how objective, reliable, and valid an assessment procedure, the results will be distorted if the behavior itself is distorted. Unfortunately, merely anticipating observation might cause non-representative behavior to appear, not to mention the observation experience itself. Below are some policies and suggestions which we have found helpful in reducing the tendency of a teacher to put on an act while being observed.

First, an observer should be in the classroom only when invited by the teacher.

Second, the invitation should be based on a plan of inquiry which was developed by the teacher and observer prior to any classroom visits. Observation should produce information which is relevant to some problem or question which is considered important to both participants. Thus, a teacher participating in an inservice training program which proposes to improve the way pupil ideas are handled during classroom discussion may be curious about this aspect of interaction before and after training. Such a question might involve creating two similar lesson plans in which a teacher would be confronted with opportunities to react to pupil ideas. One lesson would be observed before training and the second lesson after training. The plan could be embellished to provide greater insights by collecting additional data. For example, predictions about pupil perceptions, teacher perceptions, pupil attitudes and similar phenomena could be made, one lesson compared with the other. Then, instead of merely counting the incidence of constructive teacher reactions to ideas expressed by pupils, certain theories about the consequences of such teacher behavior might be investigated.

Third, the status and power difference between the observer and the teacher should be a minimum. Another teacher who is a best friend might make the most appropriate observer, providing skill in observation is present.

Fourth, the conference to provide feedback should follow a logical plan of inquiry. All the relevant data should be at hand and referred to in terms of questions to be answered and not in terms of idle curiosity. It often helps to have two or more matrices, since this facilitates the making of comparisons out of which theoretical explanation grows. A single matrix more often stimulates opinions about what is "good" and "bad," illustrated by the question, "Do you think the lesson was satisfactory?" Such questions place the observer in an awkward position, since he must express a general judgement. In this position he cannot be an equal partner in the inquiry process. On the other hand, when a hypothesis about behavior is being investigated and professional competence of the teacher is not in the foreground, the analysis of data is more systematic and unwanted defensive distractions are less likely to occur.

Fifth, the entire procedure including planning, execution, and analysis usually works more smoothly when the teacher, as well as the observer, has had approximately equal experience in observation. For example, setting up hypotheses, designing two comparable lesson plans, and knowing where to look in a matrix for the proper information are phases of the experience which should be shared by two partners who are equally competent. When the observer is more experienced and competent, the teacher defers and becomes the dependent member of the team. Both members of the team should have had previous experience in both teaching and observation.

ILLUSTRATIONS OF PRACTICAL PROCEDURES

For purposes of illustration, let us assume that the goal of inservice training is to increase the teacher's skill in making use of ideas expressed by pupils. In an article as short as this one, only four aspects of such an inservice training program will be mentioned. First, some initial performance data provides a before training performance pattern for which subsequent observation data can be compared. Second, skill training procedures can be closely correlated to observation procedures. Third, data in

TABLE 3

OBSERVATION MATRIX

CATEGORY	1	2	3	4	5	6	7	8	9	10	Total
1	-	-	-	-	-	-	-	-	-	-	-
2	-	1	1	1	2	-	-	1	5	-	11
3	-	-	5	1	4	-	-	-	-	-	10
4	-	-	-	23	2	1	-	43	1	3	76
5	-	3	1	22	80	1	2	3	3	3	117
6	-	-		1	-	-	1	3	-	-	5
7	-	-		-	2	1	1	-	-	-	4
8	-	5	-	22	19	-	-	43	7	-	98
9	-	3	3	3	7	-	-	3	32	-	51
10	-	-	-	3	1	2	-	1	1	-	8
Total	-	11	10	76	117	5	4	98	51	8	380
%	-	2.9	2.6	20.0	30.6	1.3	1.1	25.8	13.4	2.1	100
of	25.5				33.2			39.2		2.1	
Total	Teacher Total 58.7							Student Total		Silence	

I/D = 0.77 Content Cross = 70.8%

$I/D_{8,9}$ = 1.38 Steady State = 49 %

addition to interaction analysis are helpful in deciding whether or not a change in behavior is an improvement. And fourth, more advanced training designs and more complex data collection procedures will be necessary to push progress beyond the initial results. Each of these four topics will be discussed in turn.

INITIAL PERFORMANCE DATA

Initial performance data might be in the form of a short observation during a class discussion. Table 3 shows a matrix of a teacher which will now be interpreted to show how the observer and teacher make an initial diagnosis before training. The same data, of course, can serve as a before training standard in order to determine whether change has occurred.

Since the total tallies equals 380, one can estimate that the matrix represents about 19 minutes of interaction (100 tallies = 5 minutes, 20 tallies = 1 minute).

A number of percentages and other ratios, which help to form an initial picture, can be found at the bottom rows of Table 3. For example, the teacher talked 58.7 percent, the pupils 39.2 percent, and silence and confusion was 2.1 percent. The teacher was fairly directive, that is he initiated more than he responded, as shown by an I/D = 0.77 (divide all tallies in categories 1 + 2 + 3 + 4 by 5 + 6 + 7 to obtain this ratio). The highest cell frequencies are found in the steady state diagonal cells, such as the 5-5, the 8-8, and the 9-9 cells. This suggests that the teacher and pupils were able to continue a particular mode of expression once it started. Higher frequencies in these steady state cells indicate that the tempo of exchange was slower, for example, than might occur in a drill period. All silences were 3 seconds or less (note no tallies in the 10-10 cell) and most pauses followed teacher questions (N = 5 in the 4-10 cell) and teacher lecture (N = 3 in the 5-10 cell) rather than pupil statements. In six of these eight transitions, it was the teacher who broke the pause by talking (note that N = 1 in the 10-8 and 10-9 cells). This analysis of silence supports the interpretation that the teacher tended to initiate and did not permit a pupil more than 3 seconds to respond.

The matrix is sometimes more easily understood when it is translated into a flow pattern illustrated in Figure 1. In this diagram the most frequently occurring steady state cells are shown as rectangles and the size of the rectangle indicates the relative frequency of the pair. Transitions among these cells are indicated by arrows and the thickness of the arrow is roughly proportional to the frequency of these transitions. Anyone can learn to draw such a pattern flow diagram from a tabulated matrix. Begin with the highest frequency cell, in this case the

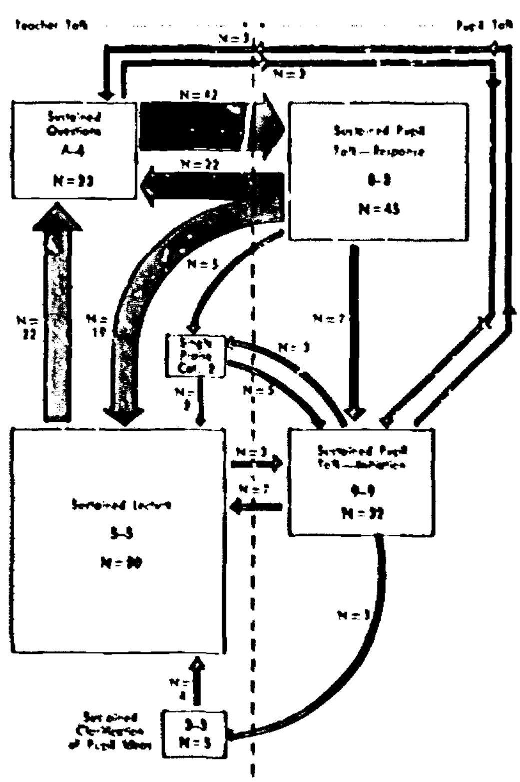


Figure 1. Initial Flow Pattern Before Training

5-5 cell with eighty tallies. Proceed across the same row to find out the next most likely occurring events and inspect columns to identify the most likely preceding events. The arrows in Table 3 illustrate the clockwise direction of the flow. Thus in both the matrix and in Figure 1, begin with the 5-5 cell, sustained lecturing. The next most likely event is that the teacher will ask a question. Next, pupils are most likely to respond to this question, see category 8. There are two events of almost equal probability following pupil response talk, one is that the teacher will ask another question and the other is that the teacher will lecture. The four transitions shown by the heavy arrows in Figure 1 account for most of the interaction: these are, 5-4, 4-8, 8-4, and 8-5. Other transitions do occur, but the transitions just listed and their associated steady state cells account for 285 tallies or about 75 percent of the entire observation.

Many things could be said about this pattern of interaction, but by far the most certain inference is that in the event that this teacher increased his use of category 3, it could be clearly seen in the matrix. The use of category 3, especially the 3-3 cell, is below average for a classroom teacher. After training along these lines a second matrix and flow pattern diagram could be made as a comparison with this initial performance.

SPECIFIC SOCIAL SKILL TRAINING EXERCISES

Specific social skill training exercises can be conducted which are closely correlated with interaction analysis. A training skill exercise might involve three or four people who take turns providing certain patterns of verbal statements while responding spontaneously to the statements of others in the group. The person being trained is designated as the actor, one other person is the observer, and the rest are reactors or foils who provide spontaneity. Different assignments are given to the actor which he then carries out by interacting with his foils. The observer keeps a record of the communication by coding within the ten categories. The assignment can be given in terms of the categories, for example, the actor is asked to produce a 4-8-4-8-4-8-4-8, etc., sequence. An actor might be asked to choose a topic and produce any sequence which is relevant to teaching and is related to the goals of the training.

The activity of a training exercise follows a model in which an actor attempts a particular pattern, his foils provide a spontaneous setting, and the observer keeps a coded record of what happens. Some training episodes may last only a few minutes, others are longer. Each episode is followed by a discussion of the performance and ways that it can be improved. While this kind of activity can easily degenerate into superficial play acting, especially when first attempted, it is also true that serious and interested participants can use this technique to practice producing certain patterns of behavior under increasing difficult circumstances and make progress in the self control of spontaneous behavior.

Learning how to make full use of the ideas expressed by pupils is a goal to which spontaneous skill training exercises can easily be adapted. It is possible to gain practice in listening carefully, in listing ideas expressed by another person, in summarizing ideas which have just been expressed, in building questions on ideas expressed by others, and to teach pupils to initiate their own steps in problem solving by showing a pupil how his ideas are related to a problem.

Skill practice sessions become more realistic when they are closely related to classroom practice sessions. After one or two initial observations an observer and teacher may discuss the relationships between the teacher's intention and the patterns which appear on several matrices. Skill practice sessions are then designed to emphasize a personal goal of the teacher based on discrepancies revealed by the matrix. If the training activities are custom built to the needs of the teacher, and if he sees them as relevant to his own professional development, then he is more likely to approach the training sessions with serious intent and a sense of optimism concerning personal development. It also helps when classroom patterns begin to change in a desired direction as a result of skill practice sessions.

Usually, the assignments of actor, foil, and observer are rotated during skill practice sessions. In this way a teacher not only becomes familiar with the coding system of interaction analysis, he also obtains some degree of observer proficiency. He also learns to accept and develop the perceptions that others have of his behavior.

The primary purpose of these training episodes is to practice producing certain patterns of statements, to translate concepts about teaching into spontaneous behavior patterns, to learn how to assess such patterns, and to discuss how these various patterns appear in classroom teaching. Many variations of training exercises can be directed toward these goals making them useful for inservice training.

DECIDING WHEN A CHANGE IS AN IMPROVEMENT

Deciding when a change is an improvement usually requires data in addition to interaction analysis. For example, merely giving a teacher instructions to "go in there and produce more threes," much like a football coach, is quite likely to increase the incidence of category three. This change may take place, however, without any insight into the teaching process and may possibly be seen as either inappropriate or not realistic by the pupils.

Utilizing the ideas expressed by pupils during classroom discourse involves several pedagogically sound principles. One such principle is that this kind of teacher behavior stimulates the perception among pupils that each pupil is free to express his ideas. Simple paper and pencil reaction sheets, to be filled out by pupils old enough to read and write, can be used to assess pupil perceptions immediately following a practice session in which the teacher tries to accept and clarify pupil ideas. When both the observer's records and the pupil's average perceptions from a paper and pencil instrument indicate greater expression and use of pupil ideas, then the additional evidence provides greater confidence that the change is an improvement.

Another pedagogical principle is that a pupil learns to cluster ideas because they are similar and then abstract the cluster with an appropriate label through directed practice in expressing his ideas. A second observer can keep an inventory of separate ideas expressed by pupils, he can note instances in which the pupils, instead of the teacher, noticed a cluster, and finally, he can record whether the teacher or the pupils supplied an appropriate label to a given cluster. This additional evidence also helps to show whether a change in teacher behavior is, or is not, an improvement.

PUSHING TRAINING BEYOND SUPERFICIAL CHANGE

Pushing training beyond superficial change usually requires special category systems of interaction analysis. For example, the division of category 3, mentioned earlier, can provide a record of different kinds of teacher reactions to the ideas expressed by pupils. A similar expansion of categories 8 and 9 will show variations in the different kinds of pupil statements. Often the expansion of teacher talk categories can provide more intricate and difficult social skill training assignments. Working back and forth between more elaborate category systems and more complex spontaneous social skill training assignments helps a teacher understand principles of pedagogy in terms of his own behavior. Unless a teacher can act out his insights about teaching, these insights are of little use for the improvement of instruction.

SUMMARY

The use of interaction analysis in an inservice training program places an emphasis on the analysis of spontaneous verbal behavior. This emphasis helps to translate ideas about teaching into classroom application.

For those who would like to try classifying verbal statements, but are inexperienced, a number of references are listed at the end of this article which include suggestions for beginning.

REFERENCES

Flanders, N. A. Intent, action and feedback: A preparation for teaching. Journal of Teacher Education, 1963, 14, 251-260.

Flanders, N. A. The role of the teacher in the classroom. Minneapolis, Minnesota: Amidon, 1963.

Flanders, N. A. Interaction analysis in the classroom: A manual for observers. (Rev. ed.) School of Education, University of Michigan, 1965.

Flanders, N. A. Teacher influence, pupil attitudes, and achievement. Cooperative Research Monograph No. 12, U. S. Office Of Education, OE-25040. Washington, D.C.: U. S. Government Printing Office, 1965.

PART V

The Current Scene

THE CURRENT scene in educational research and development is one of rapid and unprecedented change brought about largely by the advent of Federal funding. In the first paper in Part V, R. Louis Bright and Hendrik D. Gideonse trace the development of these changes with reference to involvement of the Federal government starting with the 1954 legislation. They cite the national need for the results of educational research and development as the basis for the decision of the government to provide grant dollars for developmental projects and the design of the necessary strategies to implement programs of research and development. The evolving strategies of research and development designed to improve education are delineated by the authors and include 1) curriculum development, 2) research and development centers, 3) the educational research information centers, 4) a program for training educational researchers, and 5) a system of autonomous educational laboratories. These strategies have developed rapidly and promise to provide the basis for much of the future growth of educational research and development.

The costs of educational research and development have grown from the original 1.7 million appropriated in 1957 to the present level of approximately 200 million. This represents less than one-half of one percent of the total amount expended for education by all phases of our society, a figure which is far under the standards for research and development investments in most major industries. Clearly additional growth can be anticipated. The authors of this paper propose additional major expenditures in research and development amounting to some billion dollars annually which would represent the still modest figure of approximately 2.2 percent of the total expenditures for education.

This continued growth of research and development in education, which is almost exponential, calls for careful and systematic planning. In answer to this need the Bureau of Research, U. S. Office of Education, has proposed the establishment of Educational Policy Centers to conduct studies for establishing the means and directions of educational research and development 5, 10,

20, and 30 years hence. The rationale of these policy centers is discussed by Bright and Gideonse.

The rationale and focus of a specific Research and Development Center is presented by Herbert J. Klausmeier, The Director of the Wisconsin Research and Development Center for Cognitive Learning. The Research and Development Center concept was proposed by the U. S. Office of Education in an effort to better coordinate research and development activities and to establish programatic research and development in specific areas related to the educational process. The Wisconsin Center is described in terms of eight components of operation: (1) problem area, (2) outcomes sought, (3) research and development program and strategy, (4) dissemination and adoption program and strategy, (5) staff, (6) administration and management, (7) locus of operation, and (8) source of stability of funding. Klausmeier points out that the programatic activities of the Wisconsin Center which have been brought sharply into focus now include 1) processes and conditions of learning, 2) processes and programs of instruction, 3) facilitated environments. Other programs deal with disseminotion, training activities, and technical support. Such a center concept brings to bear on problems of education the diverse talents of the interdisciplinary staff available at a major American university.

Most new and complex activities are more readily understood if represented in a model idealizing their essential components and operations. The total research and development effort in education is no exception. The model proposed by Hendrik Gideonse in the final paper recognizes the independent characteristics of the research act and the development act. In a prior model proposed by Guba and Clark the research activity precedes and feeds results into the development activity which in turn provides materials and programs for adoption. However, Gideonse proposes a model which emphasizes the nonlinearity of the research and development complex. He states, "My purpose in developing an alternative model is to create a heuristic which illustrates the essential differences between research and development activities and show how the two are—or can be—related to one another and to the operating educational system."

The model proposed emphasizes the quite different objectives and outputs of the research and development activities. Research is characterized as an activity begun with specific outcomes not being known. Contrasted to this is the development activity which is begun with the objective known or established at the onset. In the model proposed the dynamics of initiation of research and development activities are quite different, as are the outcomes or products. Research may be initiated for its own sake and pursued solely for the knowledge which it produces. Whereas development activities are initiated in response to a specific need; the specifications for the product can frequently be defined before

the onset of the development activity. However, even though the utilization of the research product is not predicted or predictable there rests on the researcher the obligation to present results in such a way so as to allow transfer to other research or development projects. In the model proposed by Gideonse the output function of the various activities is crucial. Such a model allows those concerned with planning research and development activities to more clearly see the essential characteristics of these activities.

The authors of this section describe the current scene and provide a glimpse of the future of educational research and development. The growth of the present research and development investment may indeed be a prologue to the future.

Research and Development Strategies: The Current Scene

R. Louis Bright and Hendrik D. Gideonse
United States Office of Education
Washington, D. C.

IT IS MOST stimulating to participate in a field which has generated as much excitement as educational research has in recent months. Controversy, discussion, and debate are very much the order of the day; the gradual growth of contacts with new and informative sources of data, and of faith in the research effort are developments full of promise for the future.

In agreeing to prepare this paper, we explicitly accept the notion that it makes sense to consider strategies of research and development for improving education. At the same time we do not make any claim to producing anything more than tentative steps in the direction of more adequately defined strategies for achieving mutually agreed upon objectives. For a variety of reasons which we will attempt to spell out in this paper, we have begun the process of developing and applying more explicit strategies to the research and development process. These strategies are neither rigid nor lightly held. Some notions of what we are about and why are clearly necessary; not to have them is to default a major responsibility. In this paper we try to spell out both the general outlines of the strategies we are pursuing and the procedures by which we hope to continually modify the specific substance implicit in those strategies and the actual strategies themselves.

The improvement of American education depends upon the systematic investigation of the process and the necessary conditions for learning, the development of instructional objectives, strategies, and materials based on the knowledge educators and others accumulate about the learning process, and finally on the implementation of those strategies and the use of those materials in instructional settings across the country. We take this as a fundamental article of faith, but it is a faith which has been substantiated over and over again in all fields of human endeavor whether they relate to space, health, poverty, industry or education. To claim, however, that significant and continuing advances in the process of education depend upon research and development is only the barest of opening statements. A perspective and the setting need first to be sketched in, and then a full exploration of the ramifications of that article of faith needs to be conducted.

SETTING THE PERSPECTIVE

Any consideration of the role of research and development in American education necessarily relates to the nature of the existing "system" through which we have institutionalized education in our society. With legal responsibility for education in this country vested in the states and operational responsibility for schools largely delegated to some 25,000 local educational agencies authorized by the states, the role of the Federal government has largely been confined to providing specified categorical financial support.

These different levels of responsibility are complicated by another feature of American education. Not many of the diverse institutions and agencies which serve education in this society are well-coordinated or formally related to one another. When we use the word "system" as applied to American

education, we are being generous. What we have is a de facto system which serves the broad needs of the society well, but perhaps not as well as it might if it were generously laced with efficient communication channels designed to make it an integrated, functionally related entity of many parts and many purposes.

From the point of view of the general improvement of education, it makes more sense to talk of the educational system as composed of classes of institutions and agencies. Each class embraces organizations which perform similar functions and which, if there were an organized system, would be expected to fulfill a particular role. Thus, the classes of institutions include the three levels of responsibility for education - local, state, and federal. They include professional associations and institutions for training teachers. They include the industries which supply materials and equipment to the schools. They include universities and institutes under whose aegis research on education and learning is performed. The existing "system," in other words, is largely a collection of nearly unrelated sub-systems or institutions representative of sub-systems. This feature is of no small importance in devising strategies for the general improvement of the whole.

A NATIONWIDE NEED FOR EDUCATIONAL RESEARCH AND DEVELOPMENT

The nationwide nature of the need for educational research and development is becoming increasingly apparent. Schools, districts, state education agencies, private institutions, colleges, and universities all have major needs relating to the improved fulfillment of their instructional responsibilities. While individual institutions and agencies differ, many if not a majority of the problems which confront them are similar. Serious questions of economics of scale have been answered in effect by the relatively small amounts of money which have been devoted to educational research at the local level. We have seen for decades an entirely reasonable response on the part of local and state agencies with respect to the funding of major research and development efforts. While they have no doubt been aware that sophisticated, well-conceived research programs would benefit not only themselves but also many others as well, very few have felt that they themselves could assume the financial and human burden of mounting large-scale research programs. It would make vastly more sense if the actual cost of generating those findings and materials were amortized more proportionately across all the potential beneficiaries.

The method for teaching mathematics, for example, is not likely to be much different for similar children in Maine and Arizona. On the other hand, there may be significant differences in instruction in mathematics within a state for children having different socio-economic characteristics, learning styles, or motivational problems. It would be a philanthropic district or state that felt it could mount the requisite research and development projects to serve all the immediate local needs when some other mechanisms could be developed which would distribute the cost more fairly in return for the mutual benefits that would be derived.

THE EMPLOYMENT OF NATIONAL RESOURCES

The resources for conducting educational research and development, though not as limited as some people would have us believe, are in fact unequally distributed across the country. No necessary conclusions with respect to continuing patterns of supporting educational research can be drawn from that fact, but a consideration of it is certainly one of the relevant contextual matters for the development of research and development strategies. The nationwide character of the need for research and the unequal distribution of the human resources for research reinforce one another. They highlight the breadth of the problem. In other words, so long as human resources are unequally distributed, the direct or implicit coordination of them to serve research and educational needs of high priority will constitute one of the major background considerations in research strategies.

FINANCIAL RESOURCES

Certainly one of the more important matters lending perspective to research and development (R and D) strategies for improving education is the availability of financing. The resources are limited. Money available from all sources for R and D in education has never exceeded 1 percent of the total annual expenditures on education in the nation as a whole. Estimated expenditures for this fiscal year for all levels of formal education are \$48.8 billion.[1] By contrast the funds available at the Federal level for educational research and development are somewhere between \$200-220 million.[2] Simple arithmetic reveals that this amount constitutes less than half of 1 percent of the total annual expenditure on education. We do not by any means take that as "a given" for the future, but it is most assuredly an important element in developing perspective on the problem of research strategies for the present.

THE DISTINCTION BETWEEN "RESEARCH" AND "IMPROVEMENT"

A final element for consideration in this context is our conviction that R and D in education and the change process in education are two very different functions or processes which must, however, be related intimately, sensitively, and integrally to one another. The research and development process produces the substance of improvement. It yields the knowledge about the learning process and the social setting for learning and education, and encourages the refinement of objectives and the development of materials and instructional processes which incorporate that knowledge into improved educational practices.

The implementation of those practices in the school systems across the nation - the differential accomodation of different institutions and agencies to the new roles and expectations created by the improvements - however, is a very different kind of problem and requires strategies and tactics commensurate to the task as it is revealed. The differences between the two processes and the unique

characteristics of each may very well shape significantly the strategies adopted for one or the other. The functional demands of educational research and development, for example, place certain constraints on the process of implementation. Similarly, the exigencies of the process of implementation may impinge upon the strategies for R and D which may be adopted.

RESEARCH, DEVELOPMENT, AND THE PROCESS OF IMPLEMENTATION

Recent policy-level discussions of educational research and development within agencies responsible for research funds, between such agencies and their constituencies, and within the research community itself, have made it clear that there is far from unanimous agreement on the definition of the terms commonly applied to research and development. For this reason, we believe it would be helpful to establish how we are now using various terms - not to develop for all time a thesaurus as applied to the research process, but rather to clarify our concepts and interpretations.

The research and development process involves all those activities designed to produce knowledge about the learning process (research, for example, designed to improve learning theory or to improve our understanding of the fundamental chemical, biological, and neurological processes underlying learning). It includes other kinds of activities devoted to the social context of learning (for example, the classroom or school as a social system, or the relationship among non-instructional factors such as motivation of peers, socio-economic status, educational level of the home and the direct instructional interventions of the school such as curriculum, instructional methodology, facilities and so on). It includes studies pertaining to the organization and administration of the instructional enterprise.

Research and development also includes the development of new practices, materials, and processes based on the knowledge newly created in the other kinds of studies. This involves identification of the specific improvement desired and then the systematic engineering of the improvement, using all the information derived from basic research that is relevant.

A third part of the process, closely related to development, is the demonstration of the feasibility of the new materials, processes, or organizational forms in other than the hothouse setting of the development laboratory.

Finally, a fourth element in the R and D process is the systematic dissemination of information relating to all the different stages. Information about research projects completed and underway is essential to the basic researcher. The educational developer also needs to know the latest findings about learning if he is to be able to incorporate them in his work. Researchers and educators in the schools need to know what the development people are up to. On the other hand, the kinds of problems encountered by those engaged in curriculum or hardware development may suggest research projects to the psychologists and sociologists. School people need to keep aware of new techniques and possibilities under development. Demonstrations need to be publicized, too, and the success or lack thereof provides further information to both researchers and developers. Finally, the existence of the dissemination capability itself needs to be widely publicized to insure that all those having the need can make appropriate use of it.

Applied to any given piece of knowledge or any given improvement in one or a few schools, the R and D process and the implementation of improvement process appear to be identical. The differences between the two processes begin to become visible, however, when the systematic implementation of improvements in the entire educational system is the desired objective. The systematic and continuing introduction of improvements in education through judicious innovation and change based on research requires careful attention to the different roles played by the many independent agencies and institutions in education.

The first attempt, for example, to implement a new curriculum in economics in a secondary school undertaken with the aid of the original project staff responsible for developing it can easily be viewed as the normal demonstration responsibility of the project. But if the scope of responsibility is enlarged to include developing the capability of any teaching staff anywhere in the country where a decision to implement that curriculum has been made (that decision assumes, of course, knowledge about the curriculum and a judgement as to its suitability), the problem is clearly larger than an individual project. Attention needs to be paid to the problems of developing mechanisms for the inservice and preservice training of teachers to use the new curriculum. That effort in turn may point to the particular responsibilities of teacher training institutions across the nation both to assist in the particular effort and to insure, more generally, that preservice training is undertaken in the light of the almost certain need for continuing education of teaching staff to keep up with new departures in curricular application. Thus we have moved in this one instance very quickly away from the individual school setting to the much larger question of how teacher training institutions - and maybe, therefore, also state teacher certification agencies - are to play their particular and general roles of enhancing the likelihood of the adoption of proven innovation.

The matter can be carried still further. School boards should be apprised of the need to continually re-examine curricular objectives in terms of present and future social needs; as a consequence they can be expected to play a significant part in the networks of differing responsible agencies having unique roles in the process of implementing educational improvements based on and validated by research. The responsibilities of industry in helping to develop materials, in manufacturing, and in selling them similarly need careful attention. Their responsibilities, if any, for training their customers in the use of newly developed materials also need to be considered.

Viewed on a larger scale, then, it would appear that many of the questions associated with the actual

implementation of solid innovations system-wide depend on factors more closely related to the dynamics of social change than to the actual research and development process itself. This dynamic - whether called administration, the politics of change, or continuing attention to the intricacies of the process of building conviction in and a demand for more effective materials and practices in the schools - is a necessary condition for the implementation of the products of R and D in schools. It constitutes, therefore, an absolutely essential problem for educational R and D strategists.

RESEARCH DEVELOPMENT AND DISSEMINATION STRATEGIES OF THE PAST

In describing the strategies of the past, present, and future we have elected to characterize broadly in generalized form the major outlines of attack as pursued until a very few months ago, the kinds of changes which took place and that have placed us in a crucial period of transition, and the sorts of directions we project for the future.

The availability of millions of dollars for educational research and development is a relatively new phenomenon. The recency of this growth and its absolute size give some pause for thought. A review of developments will help disclose the range of options available to us now and in the future.

The first piece of legislation specifically authorizing the United States Office of Education to support research, surveys, and demonstrations in education - The Cooperative Research Act - was passed by the Eighty-third Congress in 1954. In effect, the Congress said that research in education was important enough that the nation needed legislation to authorize Federal support for it. The Act itself, of course, represented a first step in the development of strategies for conducting R and D to improve education. Four years later, in 1958, the Congress included two titles in the National Defense Education Act (NDEA) authorizing research on the uses of new media and on modern foreign languages. These three authorizations turned out to be but a beginning. There are now a total of seven pieces of legislation including authorizations pertaining to vocational education, handicapped children and youth, libraries and information science, and an authorization under P. L. 480, a special foreign currency program based on the distribution of surplus foods.

The passage and funding of Federal legislation authorizing research in education was a major departure in and of itself. It was recognition of the conviction that Federal support needed to be directed to a problem bearing dimensions larger than those of individual states and calling upon resources, both financial and human, then currently beyond the means of state and local educational agencies.

The initial forays in support of educational research combined the identification of areas of concern and the support of technically excellent proposals submitted in response to a general call. In the very first instance, for example, proposals were requested dealing with mental retardation. As the research programs expanded in dollar size, greater proportions were administered on a nonpriority basis. Areas of importance were identified periodically and interest maintained in those areas, but the guiding principle for the support of research was the technical excellence of proposals rather than the type of research or the substantive area of concern. That principle was viable because in the early days of the research effort it was always possible to fund all the technically excellent proposals without regard to the field of study.

Over the years, substantial amounts of applied research as well as basic research were stimulated and supported. In more recent years the research program has moved into two kinds of activity: curriculum improvement and the establishment of research centers designed to explore programmatically R and D in a defined problem area in education. Even with the addition of these new thrusts the same principle mentioned before was still applicable. The field was young and growing, but it was still possible to support all of the quality proposals submitted.

THE RECENT PAST AND THE PRESENT

In 1965, several groups including staff inside the Office of Education, a Presidential Task Force, and others began to take a hard look at federally supported research in education, then nearly 9 years old. The wisdom of earlier strategies was apparent. They had been successful - in part because of the availability of dollars, in part because of the generally unsolicited mode of operation - in stimulating the growth of and interest in research and development. The field was growing; the submissions of proposals was increasing rapidly (in fact, at a faster rate than appropriations); the educational community as a whole was beginning to pay attention to the possibilities of research.

IMPORTANT RESEARCH DEVELOPMENTS

<u>Research and Development Centers</u>. The newly acquired visibility of research and development, however, led to concern over some of the apparent deficiencies of the total research effort as it had emerged to that point. First, it became apparent that the individual research projects as a whole - while the acceptable levels of quality individually - did not fit together well enough to be considered coordinated approaches to substantive problems in education. The Research and Development Centers were in part created in response to this need, but it was felt that major chunks of the research effort should be better coordinated and designed to lead to cumulative results.

Second, part of the difficulty in developing highly coordinated cumulative research efforts could be attributed to the inadequacy of the dissemination of information to the educational research community regarding the findings of completed research and the nature of current research. A strong need was felt, therefore, for the establishment of an effective research dissemination system.

Third, a careful examination of the outcomes of research in terms of service to the schools revealed that in two senses insufficient attention was being paid to the gap between the research stage and

implementation. First, the stage of development was not being supported to anything near the degree that it should. Second, very little attention was being directed to the change process whereby improvements could be actually implemented in widespread fashion.

Fourth, it was clear that the human resources available for R and D activities would need to be expanded in at least two ways. Training programs to develop new talent would be required. New agencies and institutions previously not eligible for support would need to be tapped.

Amendments to Research Acts. The response to these felt needs took the form of a series of amendments to research legislation, the creation of a series of new program efforts under the aegis of the newly acquired authority, and the passage of a new piece of legislation. The basic piece of research legislation administered by the Office of Education, The Cooperative Research Act, was amended in 1965 by Title IV of the Elementary and Secondary Education Act (ESEA). The amendments broadened the existing authority – to support research, surveys, and demonstrations in education – to include dissemination. The range of eligible institutions was expanded to virtually all kinds of public and private organizations whether profit or non-profit. Authority was included to develop programs designed to train educational research and related personnel, and to upgrade training programs. The Office was authorized to award grants as well as contracts, and the Commissioner was given authority to award funds for the construction and equipping of facilities for research and related purposes.

These amendments vastly extended the range of activities possible under the research program, and made it feasible for the research program to meet directly some of the needs identified in the reviews of the program to that time. In addition, the testimony before the Congressional committees – or the legislative history, as it is called – made it very clear that the broadened authority was to be used to bridge the gap between research and practice and to pay substantially more attention to the problems of implementing the knowledge derived from the research efforts to date and in the future. The broadened responsibilities represented by these additional authorizations have created new demands on the administration of research and at the same time have offered new tools for meeting emergent shortcomings of the research effort.

Training Program. To date, three major new program thrusts have undergone spirited development. A training program designed to expand the corps of educational researchers has been developed. This program provides both program development grants and an array of training mechanisms ranging from institutes of short and long duration through undergraduate and graduate training programs to a small postdoctoral effort (Clark, 1967).

Educational Research Information Center. The Educational Research Information Center (ERIC) has been established with Central ERIC, twelve clearinghouses, and more to come. ERIC is designed to provide to researchers and practitioners alike an information storage and retrieval mechanism to make available instantly in easily accessible indexed form all the research and related data relevant to a particular problem. Research in Education, the monthly publication of the ERIC system, has now published several issues and already the usefulness of this effort has begun to prove itself.

Educational Laboratories. The third and largest development in the research program has awakened the interest, excitement, and the enthusiasm of the entire educational community from local schools, state agencies, and teacher training institutions to industry, scholars from the arts and sciences, and the lay public. We refer to the National Program of Educational Laboratories. Drawing from resources in regions extending across the nation, new institutions called educational laboratories have been created to bridge the gap between research and practice. These institutions were created by representatives of all the many agencies and institutions which play different roles in the implementation process. They are reflections of the conviction that it is not enough to do research; that research must be followed up by development project which, having established the desired objectives – whether curricular, instructional, organizational, professional, or technical – then move to the development of solutions drawing upon the best that research has to offer. The laboratories have also been charged with the responsibility for active dissemination campaigns based on the successful development projects they and others engage in.

We have encouraged the laboratories to conceive their responsibilities broadly – encompassing much more than merely the distribution of information, although that is a critically important function among many. Clearly, one of the important steps in the diffusion of research-based improvement throughout the educational system is the establishment of demonstrations. First time demonstrations of feasibility will be the direct responsibility of the laboratories; the more widespread diffusion of the successful innovations will depend on the degree to which information about the innovation is distributed to various parts of the systems, the degree to which the innovation recommends itself to professionals, and the degree to which credible demonstrations of the new practice or curriculum are mounted in schools. This last part of the diffusion process – the establishment of real-life (not hothouse) demonstrations of the innovation without the intervention of the inventor – represents a place where the labs can work with the state education agencies and local school districts to make use of Title III of ESEA. Such demonstration efforts must be coordinated and well-conceived. Laboratories can help only in their areas of competence and program thrust and only by cooperating closely with state education agencies and the local school districts.

The laboratories were conceived in such a way as to involve in their government and their operation the many types of different responsibilities and resources that combine to form the educational system as we know it. The hope is that these new institutions – responsible for their own program development and implementation and knit closely together to form a network – will, as their resources permit, at one

and the same time engage in 1) major efforts to develop new materials, practices and organizations using the outcomes of research; and 2) by utilizing the expertise brought to them by the involvement of different agencies and institutions, pursue courses of action which help to hasten the process of improvement once tested innovations are available.

The strategy for these three new program thrusts is straightforward. If there is to be an expansion in research and development, people will need to be trained or recruited from new areas to fill the demand. In order to build structures of knowledge and cumulative improvement, dissemination networks and the material which moves through them will need to be better systematized and improved. Finally, to fill the gap between research and practice, a new autonomous institution drawing institutionally and representationally on many resources in the education system has been created.

TITLE III OF ESEA

In addition to the new authorizations for research and the programs mounted as a result of them, the Eighty-ninth Congress passed Title III of the ESEA. This program, authorizing support for projects submitted by local educational agencies designed to supplement existing school programs or to serve as models for existing school programs, was an innovation in itself. When (former) Commissioner Keppel testified before the House Subcommittee on the Elementary and Secondary bill, he told the Congressmen that he viewed Title III and Title IV together. Title IV was to be the means whereby the substance of educational improvement would receive increased impetus and attention; Title III would be the means by which local schools could inaugurate the kind of credible real-life demonstrations which would be convincing to their counterparts and become thereby one of the moving forces for the widespread adoption of tested innovation. Those responsible for Title III estimate that this past fiscal year fully 60 percent of the money went to support projects relating to the functions of development and demonstration. This year the program expects to do at least as well. Most important, for the first time local educational agencies were provided with the funds and the encouragement to experiment with new ideas.

CURRENT DEMANDS AND STRATEGIES FOR THE FUTURE

A thorough consideration of the present status of educational research on the basis of five variables provides important clues to the directions future strategies may well take. Those five variables are discussed briefly in the following paragraphs.

VARIABLES RELATED TO STRATEGIES

The Research Function. The research function has been charged broadly with 1) the basic research responsibilities relating to improving our understanding of the nature of learning and the conditions for it, 2) developing materials, practices, processes, and institutions designed to improve instructional practices, and 3) engaging in a range of activities running from demonstration to dissemination to training, in order to ensure that improvements find their way to implementation in operational settings.

The Financial Demand. The financial demand, represented by the broad responsibilities indicated above, is difficult to estimate with precision, but its dimensions can be sketched. Recent experience accumulated by the National Science Foundation, the Office of Education, and the Office of Economic Opportunity has shown that full-scale multi-media curriculum development can range in cost from $2 million to $6 million per course; even this expense is engendered by building curriculums for schools as they are presently structured. The costs for similar development for newly organized institutions might well be more. In any case, however, it does not take very complicated arithmetic to multiply the above figure by the number of possible course experiences that might be developed for preschool, elementary, secondary, community college, undergraduate, graduate, and adult education. If one assumes, as we think is reasonable, that such curricula probably ought to be redeveloped every 5 or 7 years it is not unreasonable to project an annual expenditure of a half billion dollars just for curriculum, hardware, and organizational development alone. The research effort to support that kind of development effort will add another $150 million to $200 million to that figure annually. The cost of demonstration must also be added – perhaps as much as $300 million or $400 million annually. A rough rule of thumb for dissemination activities indicates that approximately 5 to 10 percent of the research effort ought to be spent in that area. We can add conservatively for that purpose another $50 million. The total is slightly more than a billion dollars annually, an amount which – if added to the current estimated expenditures of educational institutions for this fiscal year – would constitute just 2.2 percent of the total. That rate of expenditure for research and development approximates the minimum percentage devoted to R and D today by American industry. It strikes us as reasonable that the education industry, with estimated expenditures of $48.8 billion in fiscal 1967, should devote a similar proportion of effort to research and development in areas relevant to its missions.

The Dollars Available. The actual dollars available for the broad range of activities associated with the research effort have grown dramatically. They are by no means equivalent to the demands briefly sketched above. But major funding for research and related activities is still in its infancy. Significant educational research funding began only in 1957 at a combined USOE/NSF level of $1.7 million and has grown this year to an annual level of $200 to $220 million including USOE research appropriations, the NSF Course Content Improvement Activities, other miscellaneous agency expenditures, and Title III of ESEA (assuming that the 1966 proportion – 60 percent – of the Title III, ESEA, activities going to development and demonstration holds for 1967).[3]

This level of funding, matched with the estimate of the conceivable demand for funds for research and related activities, underscores the critical nature of the allocation of financial resources.

Manpower Resources. The problem of manpower resources is of crucial importance and is not easily resolved. A direct response, although a long-range one, is to train a new corps of professionals having the kinds of skills required to perform the entire range of research functions. There are other possible approaches, however. There already are – in other disciplines and in other kinds of institutions – individuals skilled in the research and development environment, and who could probably be persuaded to turn their attention to the field of education. These considerations were in no small measure part of the strategy behind broadening the list of eligible organizations to participate in R and D. They also played a major role in the early decision in the administration of research to secure the active involvement and participation of arts and sciences disciplines in educational research, a strategy that has continued until now and will be maintained in the future.

The combination of approaches adopted for the provision of adequate manpower supplies can be expected, then, to have considerable bearing on the possibilities for growth (see Clark, 1965; Guba and Elam, 1965; Hopkins, 1967).

Successses of Past and Current Efforts. The successes of educational research and development, the growing visibility of operations in this area, and the increasing conviction that research will in the long run provide the basis for substantive improvements in education have created new demands upon the human and financial resources available to the research enterprise. As soon as the demand outstripped the resources, something that happened only very recently, the entire picture of support necessarily was altered. It has now become necessary to allocate limited resources in the face of demands which exceed those resources.

ADDITIONAL STRATEGIES ADOPTED

The five considerations discussed above have led to the adoption of additional strategies. Broadly speaking those strategies embrace the three principal functions discussed below.

Careful Long-Range Planning. Now that the demand has clearly out-stripped the supply of resources, particularly in the financial area, it is critically important that the allocation of research dollars to various research functions and to particular topics or areas of development be accomplished in a careful and logical manner. This must be done in ways to ensure that both short-term and long-term interests are met, and to maintain and strengthen the existing political structure of education as it has developed over the years.

The careful long-range planning of those resources is much more likely to produce optimum levels of support for all the different functions and areas for research and development than is a totally non-directive approach embracing the competitive selection of technically excellent but unsolicited proposals for support. Such planning will need to take into consideration the instructional and administrative needs of the different educational levels and the many kinds of agencies, institutions and organizations in the field of education, industry, the professions, and government. Attention will have to be given to the needs of different target populations, including the general student, the handicapped, the disadvantaged, and the gifted. This implies, of course, that it will be necessary for us to develop much more effective mechanisms for planning research and related activities in order to ensure that all relevant data are canvassed and that maximum flexibility is built into the planning and administrative policies devolving from that planning. We must be careful, for example, that we continue to pursue the wisdom of earlier efforts in support of research by maintaining a significant portion of the research budget for unsolicited efforts. This is only one element of the larger planning picture, however, although a very important one.

As part of the attempt to improve the data on the basis of which research and other educational planning is now done, the Bureau of Research has recently called for proposals to begin pilot projects we hope will ultimately lead to the establishment of educational policy centers. These centers would, on a continuing basis, conduct the kinds of studies which would enable all of us to have better ideas, of the 5-, 10-, 20-, and 30-year picture for education. How should or might the schools relate to the society? What should the range of curricular objectives be? What kinds of sources might exist for financing education? How might schooling be structured in the society in the years to come? What technologies are likely to be available? These centers will not be engaged in predicting what the future will be. Instead, they will try to project the many alternatives available to us, cost them out, examine the consequences, and thus provide much better data to policy planners than are now available. They will provide a new kind of information and analysis looking to the future in a systems-oriented interdisciplinary way, supplementing the improved present-oriented data collection which must also accompany the planning process (see Goodlad, 1967).

The Outputs of Research Process. Our concern for the outputs of the research process is pragmatic in part, but it is tied closely to our conception of the mission of the research effort in education in general and of USOE's research program in particular. We have stated our faith that the improvement of education depends in the long run on the systematic prosecution of a well-conceived R and D effort. The improvement of education is the fundamental thrust of the mission; that criterion must be the continuing evaluative guide we employ in judging the success of our efforts.

While it is true that the justification for supporting research and related activities in education with public money can only be the eventual and significant betterment of the educational system as a whole, it is also true, happily, that the more we can demonstrate that research has affected school practices in a positive way, the more likely it will be that greater support for the entire research process will be forthcoming. Thus we believe that such objectives are not only good public policy, but good politics as well.

There can be no denying the growth and the strength of the interest in results. The Congress is concerned. The Executive Branch is concerned. Local school people have made their interest felt in a number of ways, particularly in the development of the programs for the educational laboratories.

We believe, however, that a strong and continuing tension exists between the conduct of research designed to increase our knowledge of basic learning processes, for example, and the pursuit of development projects designed to yield specific outcomes for instructional use. The demand for results has raised the appropriate and throughly justifiable concern of the academic and scholarly research community that such demands may tend to compromise the long-range efforts which are so badly needed. The continuing tension between short-term and long-term requirements – between today's youngsters and tomorrow's – is real and constitutes one of the continuing nightmares of the research administrator. This problem is the responsibility of the administrator of research, and it must be squarely met. We believe that it is important that the two functions not be confused, but that they both be served with sophistication and energy.

The importance of concentrating on the improvement of schools, and the significance of development projects in that effort, signals an increasing attention to such activities as curriculum development and explicit attempts to improve instructional practices. We are increasingly persuaded, for example, that prudent management of the limited resources available at the present time makes it necessary for us to adopt a research strategy that relates applied research projects closely to identified development efforts. We would hasten to add that the preservation of a significant portion of funds for unsolicited proposals will still permit the support of some applied research unrelated to development since good ideas will also be generated independently of development projects or research planning groups and should be supported on their own merits.

We anticipate, then, still further increase in attention to development kinds of tasks designed to produce substantial, measurable, and cumulative improvements in the nation's schools.

Continued Expansion of Research and Related Activities. There is no question in our minds, as all the preceeding analysis should clearly indicate, that research and related activities must continue to expand in rapid but judicious ways. By directing our attention to the development of coordinated strategies for the support of research with the implication that greater attention will be paid to the outcomes of the research effort, we believe that constituencies in support of educational research will grow and the dollars available for the function will increase to what we believe are the regular minimal levels we suggested earlier.

SUMMARY

In this paper we have sketched out the two parts of the change process: research and development, on the one hand, and implementation, on the other. We have suggested that the principal purpose of the research effort ought to be viewed as the improvement of education in the schools. We have indicated the national need for research, and the importance of drawing on national resources to meet that need. We have described the discrepancy between the potential size of the need in dollar terms and the actual fiscal resources available. We have suggested that these conditions lead us to adopt, in effect, multiple strategies for R and D. These include: 1) better planning of the research effort; 2) more attention to the training of research and related personnel; 3) the establishment of new institutions – the laboratories – and new programs – Title III – to greatly increase the opportunities and responsibilities of local school officials in the research and implementation process; 4) the continuation of significant proportions of unsolicited research and development funds; and 5) increased attention to development efforts designed to yield better materials, equipment, and instructional practices for the schools. All of these strategies will, we hope, lead to increased understanding of the ultimate impact of the research effort, help to bring about substantial increases in the amount of support committed to research and development in education, and will provide the foundation for the continuing improvement and self-renewal of our educational system.

NOTE: Dr. R. Louis Bright is no longer with the United States Office of Education. He is presently located at Baylor University in Waco, Texas where he is a professor.

FOOTNOTES

1. Digest of Educational Statistics, 1966, U. S. Government Printing Office, Washington, D. C., 1966, p. 17.

2. The imprecision in the estimate derives from a number of considerations. The research authorizations of USOE in FY 1967 total $99 million, but $6 million are for training and another $14 million are for construction. The National Science Foundation spends close to $17 million a year on curriculum development, and almost certainly some of the curriculum development efforts of the Office of Economic Opportunity will be used by schools. Similarly, the National Institute of Mental Health and National Institute of Child Health and Human Development also spend a considerable amount on research related to areas of interest to learning and education. Title III of ESEA, if past percentages hold, could well spend $81 million on development and demonstration activities. Part of the development of strategies for educational R and D is the coordination of these several sources of research and research related funds.

3. The actual figures by program and year are as follows: (See chart on following page.)

	USOE *Research*	NSF Course *Content*	Title III, ESEA *Dev. and Dem.*
1954	$.....	$ 1,725	$.....
1955	.....	35,000	.....
1956	.....	18,000	.....
1957	1,020,000	650,140	.....
1958	2,300,000	750,310	.....
1959	8,716,000	6,180,485	.....
1960	10,350,000	6,302,055	.....
1961	10,117,750	6,167,740	.....
1962	11,770,000	9,389,948	.....
1963	14,188,400	12,620,771	.....
1964	19,820,000	14,157,650	.....
1965	37,703,000	14,889,081	.....
1966	100,141,241	16,393,383	45,400,000
1967	99,600,000	17,000,000 (est.)	81,000,000 (est.)

REFERENCES

Clark, D. L. Educational research and development: The next decade. In E. L. Morphet and C. O. Ryan (Editors), Implications for education of prospective changes in society. Reports Prepared for the Second AREA Conference. Denver, Colorado: Bradford-Robinson, 1967. Pp. 156-175.

Goodlad, J. I. The educational program to 1980 and beyond. In E. L. Morphet and C. O. Ryan (Editors), Implications for education of prospective changes in society. Reports prepared for the Second AREA Conference. Denver, Colorado: Bradford-Robinson, 1967. Pp. 47-60.

Guba, E., and Elam, S. (Editors) The training and nurture of educational researchers. Sixth Annual Phi Delta Kappa Symposium on Educational Research. Bloomington, Indiana: Phi Delta Kappa, 1965.

Hopkins, J. E. Scope of the demand for educational research and research related persons. Paper presented at the meeting of the American Educational Research Association, New York, February 1967.

The Wisconsin Research and Development Center for Cognitive Learning

Herbert J. Klausmeier
The University of Wisconsin
Madison, Wisconsin

THE AGREEMENT between The University of Wisconsin and the United States Office of Education (USOE) establishing the Wisconsin Research and Development Center for Cognitive Learning was signed on August 6, 1964. The high expectations held for the Center were expressed by the principals to the signing in a number of ways.

University President Fred Harvey Harrington remarked that the signing denoted "... an important day in the history of the University. This agreement marks the culmination of years of effort devoted to the improvement of the research capabilities of the School of Education, and opens the way to still greater developments."

Howard F. Hjelm, then director of basic research for USOE, described the agreement as following the new pattern in educational research of full institutional commitment. He stated: "The University has pledged itself to fully investigate the problem of learning and to disseminate research findings in a way which will bring about definite changes in school practices."

Hjelm found the University well suited for the commitment it had accepted. "The University was chosen because of the exceptional quality of its educational researchers, because of its excellent relationship with local schools and with the State Department of Public Instruction, and because of the wholehearted dedication to the project existing in Madison."

Center Director Herbert J. Klausmeier stated the primary focus of the Center: "Research emphases will be on learning by children and youth in normal school situations. Especially stressed will be the learning of concepts and problem-solving techniques in mathematics and other basic subjects."

The Wisconsin R and D Center was established under provisions of the Cooperative Research Program of USOE. In 1964 the Cooperative Research Program included six major areas: 1) basic and applied research, 2) demonstration, 3) curriculum improvement, 4) small contract, 5) research and development centers, and 6) developmental activities. Further description of the research and development centers was given in OE Publication 12017, 1963, as follows:

> Research and development centers are designed to concentrate human and financial resources on a particular problem area in education over an extended period of time in an attempt to make a significant contribution toward an understanding of, and an improvement of educational practice in, the problem area. More specifically, the personnel of a center will:
>
> 1. Conduct basic and applied research studies, both of the laboratory and field type.
> 2. Conduct development activities designed to translate systematically research findings into educational materials or procedures, and field test the developed products.
> 3. Demonstrate and disseminate information about the new programs or procedures which emerge from the research and development efforts. These activities may

include demonstrations in a natural, or operational, setting; the preparation of films, tapes, displays, publications, and lectures; and the participation in symposia and conferences.

4. Provide nationwide leadership in the chosen problem area.

The preceding ideas about the programmatic research, development, and dissemination thrust of an R and D Center in relation to a significant problem area in education followed logically from experience with other programs of the Cooperative Research Program of the Office of Education, the first of which, basic and applied research, started on a modest scale in 1956. Also, the research and development strategy had demonstrated a high yield in selected aspects of industry, agriculture, medicine, and space exploration. A large segment of the American public desired improved educational practices. Relevant USOE, university, and other concerned persons throughout the nation felt that a program of research and development centers would be equally effective in improving education.

An R and D Center, designed to improve educational practices through programmatic research and development, might be considered as a concept. As we know, concepts are abstractions of realities. A concept is comprised of the attributes, or properties, of things and events, the attributes being joined by rules, or ways of combining attributes systematically. An R and D Center then may be understood in relation to various attributes and how they are combined to form a whole. The writer hypothesizes eight crucial attributes of an R and D Center as follows: problem area, outcomes sought, research and development program and strategy, dissemination and adoption program and strategy, staff, administration and management, locus of operation, and source and stability of funding. The preceding attributes may be considered as the major components of a system, or instrumentality, designed to achieve certain general objectives implied by the problem area. Thus, a systems approach is properly followed in organizing and managing the Center operations and programs. The remainder of this chapter is devoted to a discussion of the various components of the system, their relationships, and the outcomes attained thus far by the Center staff.

Before further description of the Wisconsin R and D Center, it should be recalled that the Elementary and Secondary Education Act of 1965 as noted by Bright and Gideonse in the previous chapter, provided for the establishment of regional educational laboratories and Title III demonstration and adoption centers in local schools. These two programs of USOE are much larger than the current R and D Center Program. As instrumentalities for improving educational practices, the writer hypothesizes that they, too, can be described and understood in terms of the preceding eight attributes of a center. It is not the purpose of this chapter, however, to point to the similarities and differences among the various research, development, and demonstration instrumentalities of the USOE but to describe the R and D Center of Wisconsin as an example of a significant contributor to the extension of knowledge about cognitive learning and the improvement of related educational practices through research and development.

PROBLEM AREA

The problem area of the Center is the improvement of education through a better understanding of cognitive learning. The importance of this problem area may be inferred from five propositions regarding learning and instruction.

The first proposition is that concepts provide much of the basic material of thinking. Concepts possessed by the individual enable him to interpret the physical and social world and to make appropriate responses. Concepts once learned then serve as symbolic mediators between sensory input and overt behaviors. Without concepts with which to think, man, like lower-form animals, would be limited mainly to dealing with sensorimotor and perceptual representations of reality, closely tied to immediate sensory experiences.

A second proposition is that concepts and cognitive skills comprise the major outcomes of learning in various subject fields. Sentence, noun group, and transforms represent four concepts in English syntax. Set, number, and fraction represent concepts in mathematics. In order to learn the concepts of a discipline, a student must acquire cognitive skills. In mathematics these skills are designated by such terms as arithmetic comprehension and problem solving. The identification of concepts and their arrangement in a hierarchical order and the identification of cognitive skills are principal concerns of scholars and curriculum workers in the various subject fields.

A third proposition, related to the preceding ones, is that the acquisition of cognitive skills enables man not only to acquire the knowledge of others but also to generate new knowledge. Cognitive processes, such as cognizing, hypothesizing, and evaluating information, are learned. These and other processes, combined with cognitive contents, are designated cognitive abilities or skills. When the child has learned the cognitive skills of reading, he can acquire knowledge in the absence of other persons. When he has mastered mathematical concepts and computational skills, he can find solutions to problems not encountered in the same form by any other person.

A fourth proposition is that efficient learning of concepts and cognitive abilities is related to conditions within the learner and conditions within the situation. Two conditions within the learner associated with efficiency of learning are attention, related to motivation, and level of cognitive functioning. In school settings, the preceding conditions related to the student and other conditions related to the content and sequence of instruction, quality and availability of equipment and materials, and quality and methods of instruction are paramount. More efficient learning can be attained through better control of these conditions.

A final proposition is that knowledge about cognitive learning, concepts, cognitive skills, and conditions of learning generated by scholars requires

validation in school settings. Although the study of cognitive learning is in its infancy, sufficient knowledge has been generated to develop instructional theory regarding the teaching and learning of concepts. The validity of the theories must be tested in school settings. Similarly, instructional materials for children and outlines of concepts for teachers must also be tested for usefulness and validity in school settings.

PROGRAMS, STRATEGIES, AND OUTCOMES

Strategies best suited to achieve the Center goals include the differentiation of the problem area into research and development programs and other programs. The three research and development programs are titled Processes and Conditions of Learning, Processes and Programs of Instruction, and Facilitative Environments; the other programs deal with Dissemination, Training Activities, and Technical Support. These programmatic activities are designed to concentrate and utilize human and monetary resources most effectively over an extended period of time. Each research and development program is characterized by a sufficiently sharp focus to assure its identity and by a relatedness of content and methodology that facilitates the free flow of information and personnel among programs. Figure 1 shows the major operation of the Center and the program components. These operations and components are now described in some detail, with attention first directed to the three R and D programs and later to the other programs. The objectives of each program are given and outcomes attained through 1968 are summarized.

PROGRAM 1 -- CONDITIONS AND PROCESSES OF LEARNING

To increase our understanding of conditions and processes of cognitive learning, basic research is needed. To test in school settings the knowledge from basic research, relevant models and systems must be developed. To meet these requirements, the research and development activities comprising Program 1 are directed toward: a) generating new knowledge about concept learning and cognitive skills by conducting laboratory-type experiments or research within the school setting; b) synthesizing existing knowledge and developing general taxonomies, models, or theories of cognitive learning; and c) developing educational materials suggested by the prior activities.

Substantive information from the basic studies in learning feeds directly into the area of instructional improvement, thus providing from the behavioral sciences a foundation for the development of instructional programs. Also, methods developed in the behavioral sciences are utilized in the research and development activities of Program 2. At the same time, the behavioral scientist in Program 1, when dealing with the learning of content in a subject field, utilizes personnel from Program 2 and other sources who are expert in the subject field, familiar with instructional processes, or both.

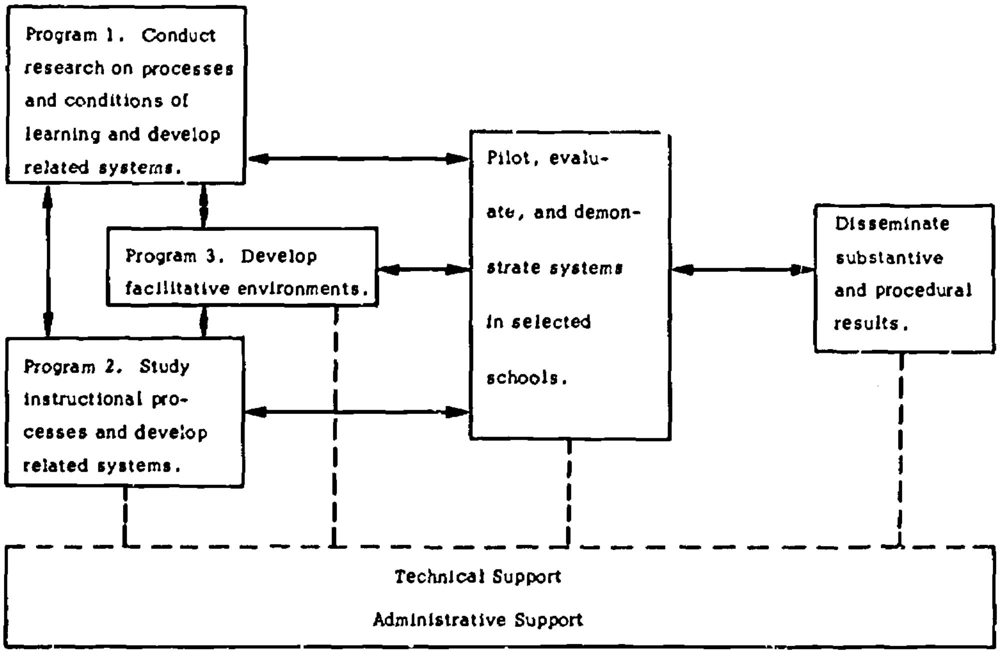
Basic research in cognitive learning has been conducted by the Center staff and reported regarding concept learning, creativity, prereading skills, rule learning, the shaping of complex human behavior, group pressures on individual performance, and the transmission properties of various media. Relevant results from these studies have gone into the development of instructional programs. In addition, a system of motivation was formulated and put into operation in selected elementary schools during 1967-68. Creativity training materials and procedures were prepared and received initial testing in selected schools in 1967-68. Procedures for teaching severely retarded readers were developed and field tested, starting in 1965, and were used operationally in selected schools in 1967-68. Other basic research dealing with concept learning and cognitive skills is continuing and is being applied to improvement of instructional programs in arithmetic, English, reading, and science.

PROGRAM 2 -- PROCESSES AND PROGRAMS OF INSTRUCTION

What is an appropriate unit to deal with when attempting to improve instruction, or stated differently, to achieve greater efficiency of student learning? The Center staff and school people decided that it is each building or an integrated set of buildings which the same children attend rather than individual children, separate classrooms within a building, or all the buildings of a local school system. Within each building, the instructional and other staff must deal with all the components of an instructional system. Similarly, the R and D staff, particularly those in Program 2, are concerned with all the components of an entire instructional system, although they may give much more time and effort to developing and testing one component rather than to another. Figure 2 outlines the major components of a total instructional system that are relevant to a school building. The order from top to bottom also indicates the general sequence in which the staff of the R and D Center and schools move in developing a prototypic system, such as in reading, mathematics, science, and English. Variables associated with each component require research. Thus, research on instructional processes and the development of instructional programs are essential in this program. The specific objectives of Program 2 are: a) to establish a rationale and strategy for developing instructional systems in the cognitive domain; b) to identify by careful synthesis and further research sequences of concepts and cognitive skills within and across disciplines; c) to develop assessment procedures and materials for the concepts and skills identified above; d) to identify existing materials or develop new instructional materials associated with the concepts and cognitive skills; and e) to generate new knowledge through research about instructional procedures including motivation, individualization, classroom management, and organization of instruction.

Executing the activities of this program requires specialists in a subject field, in instructional methods, or both, with assistance from behavioral scientists and communication experts. Each project in this program has at least one staff member whose speciality is cognitive learning. In addition, the program utilizes the many resources outside the Center including the University educational TV station (WHA-TV) and Educational Testing Service.

Figure 1: Major operations and components of the Wisconsin R and D Center for Cognitive Learning



The largest instructional program developed by the Center staff through 1968 was Patterns in Arithmetic, a program comprised of TV video tapes, student lessons, and teacher notes. It is designed as a complete program of materials for children and also for simultaneous in-service education of teachers. As of 1967-68 the materials for Grades 1-4 had been developed and initially tested and were undergoing further field testing in many schools of Alabama and Wisconsin. The program for Grade 5 was being completed. The development of a prototypic instructional program, not using TV procedures and materials, dealing with concepts in mathematics, Kindergarten-Grade 6, was started in 1967-68.

Fundamental research was conducted and reported on the learning and teaching of reading, 1966-68, and is continuing. An outline of major content, behavioral objectives, and related appraisal procedures was developed, Kindergarten-Grade 6, and relevant instructional materials and procedures were identified. The entire prototypic instructional system was piloted in selected elementary schools, 1967-68, as cooperative school and Center activities, with substantial support of ESEA Title III and Title I funds.

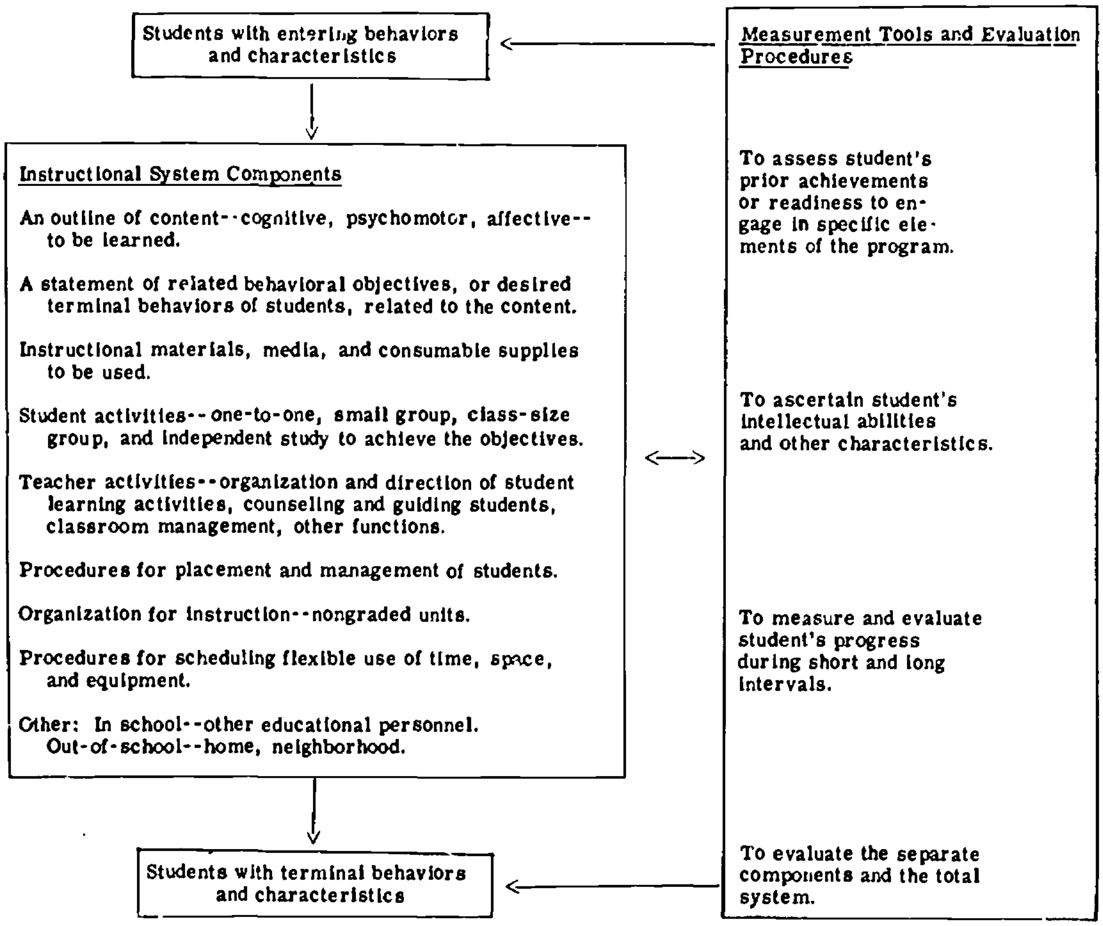
An outline of concepts related to the particle nature of matter and environmental conservation, appropriate for the elementary-school science program, was developed, 1965-68. Research on instructional methods was reported and other research is continuing. Preliminary instructional materials and methods appropriate to the concepts were developed during 1967-68 and initially tested. Related research and the development of scientific equipment and materials is continuing.

A prototypic program of instruction in English in the junior high school was under development, starting in 1966. An outline of selected concepts of English syntax and a related set of programmed instructional materials to teach the basic concepts were developed and field tested during 1966-67. Further testing of the programmed material, and implementation of the entire curriculum, was started in selected junior high schools during the 1967-68 school year and is continuing. Basic research on processes and conditions of learning is continuing.

An outline of concepts in verbal argument was developed, 1965-67, and related tests of critical thinking were also developed and initially validated. Programmed material was prepared during 1967-68 to teach these concepts. Further research on the program and validation of the tests continued, 1967-68.

Two other sets of activities have high priority for initiation in subsequent years. Research and related development regarding social studies in the elementary school are projected. Greatly increased use of the computer in research and in the management of individually guided education is the second field to be given increased emphasis.

Figure 2: Major components of an instructional system



PROGRAM 3 -- FACILITATIVE ENVIRONMENTS

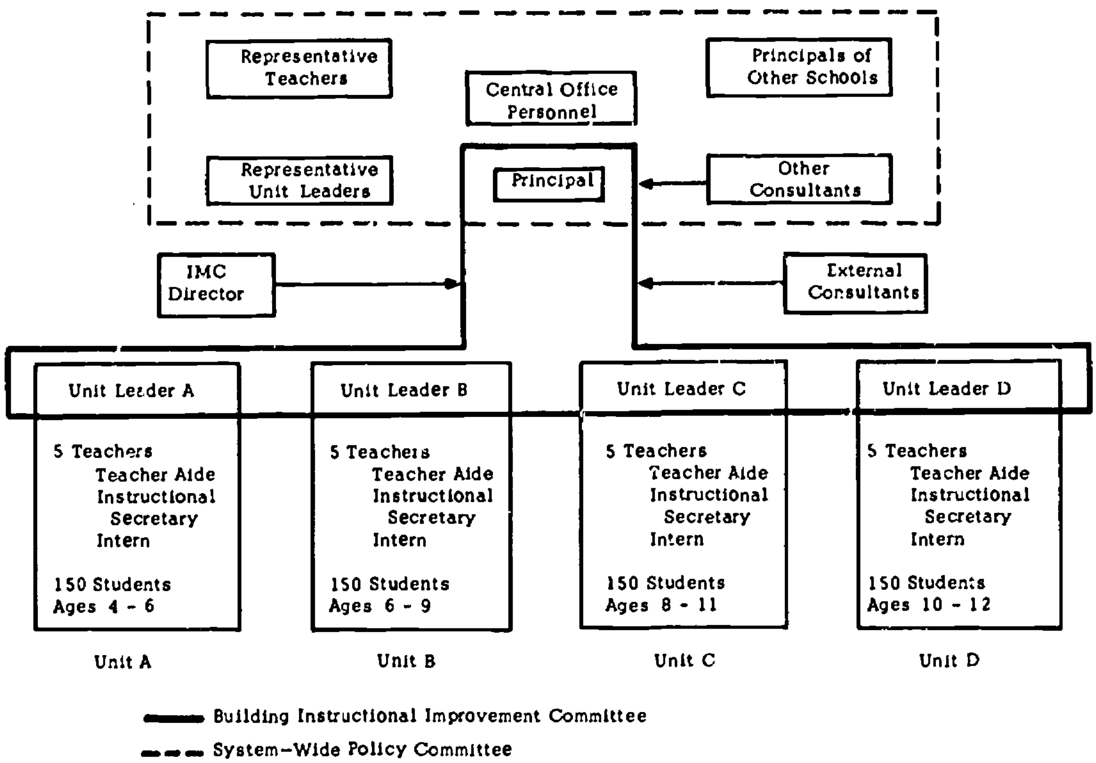
Achieving improved educational practices requires facilitative environments in selected local school buildings, in total school systems, and among various agencies. Thus, the research and development activities within this program are designed a) to develop and test new organizations and relationships among staff that facilitate research and development activities on cognitive learning in schools, and b) to develop and test the effectiveness of the means whereby schools select, introduce, and utilize the results of research and development. The same schools that facilitate student learning also serve as demonstration centers for entire school systems and the R and D Center.

In seven school systems of Wisconsin environments were established, 1965-68, designed to facilitate the research and development activities of the Center and the learning of children through a system of individually guided education. A new organization for instruction that replaces the self-contained classroom by instruction and research units and an instructional decision-making committee is an integral component of the facilitative environment of the Multiunit School. The organization is shown in Figure 3. The characteristics of individually guided education and of the organization which facilitates its execution merit further attention.

The Multiunit School organization includes both a formal organizational structure and a procedural style consisting of several essential components as shown in Figure 3. The organizational hierarchy of the Multiunit School consists of groups at three distinct levels of operation.

At the classroom level are the Instructional and Research (I and R) units. Each I and R unit has a unit leader or professional teacher, two or more regular staff teachers, one or more aides or secretaries, and in some cases an intern. The number

Figure 3: Organizational chart of a multiunit school of 600 students



of instructional staff varies with the size of the unit. The intern assumes instructional responsibilities and does not perform routine and clerical duties. Each unit is charged with planning and conducting the total school experience of about 150 students in the prototype shown in Figure 3.

At a second level of organization, the principal and the unit leaders constitute the permanent Instructional Improvement Committee of the building. The principal chairs the group, which meets weekly, more often if necessary. This committee may bring in consultants from the central office, the state education agency, or other agencies. Instructional decisions made by the committee are executed in the units.

At the third organizational level is the System-wide Policy Committee. Chaired by the superintendent or his designee, this committee includes representative principals, unit leaders, teachers, consultants, and other relevant central office staff. The personnel of this Committee have the capabilities and specialization necessary to assist the staff of the Multiunit School. The Committee meets less frequently than either of the other groups, but its operation is one key to the success of the Multiunit School.

The organizational pattern of the Multiunit School thus differs from that of the traditional, self-contained classroom school in several ways. First, in the Multiunit School personnel work in units or committees, rather than in isolation as is the case in the traditional school. Second, three new roles are added: unit leader, teacher aide, and instructional secretary. Finally, the addition of new roles and the use of personnel in groups rather than alone results in considerable redefinition of the familiar roles of principal, teacher, and consultant.

In individually guided education in the Multiunit School, the Instructional Improvement Committee determines the objectives for the particular school building, taking into account system-wide and state regulations. These are broad institutional objectives for the school building. The staff of each nongraded Instruction and Research Unit, the replacement for the graded homeroom or self-contained classroom, then decides the objectives for each child in the unit. While the unit leader takes the initiative here, each unit teacher also participates. Assessment of the child's characteristics is through observation and by means of locally constructed and standardized instruments of various types. On the basis of the assessment each child is then placed in one-to-one, small group, class-size group, and unit-size group activities. Instruction employing tutorial work and computer-assisted instruction are examples of one-to-one activities. Activities in

small groups of two to fifteen are organized to attain socializing and also skill objectives. In connection with skill objectives, 150 children in a unit might be placed in fifteen small groups for most of their mathematics instruction and then regrouped in another fifteen groups for part of their reading instruction. Class-size or homeroom activities are used for achieving any objective where heterogeneity is desired. Large unit groups of 40 to 150 are formed mainly for giving information to the total group or for independent study. The information is given by a teacher, television, sound motion picture, or by other means. Some music and physical education activities are conducted in groups larger than the usual class size. Independent study is carried out in small groups, class-size groups, and large groups.

During 1967-68 approximately 4,600 students in the elementary school, 1,600 in the junior high school, and 400 in the senior high school were enrolled in R and I Units in Wisconsin schools. In these schools, children, educational practitioners, behavioral scientists, and other staff members of the Center are brought together in a close working relationship. Thus the selected schools of Wisconsin providing the facilitative environment comprise a primary link between improved school practices and Center research and development.

DISSEMINATION

The R and D Center is committed to disseminate the products of its research and development activities to appropriate personnel by various media and also to install them in selected schools where others may directly observe the related educational practices. So that the Center may fulfill its obligation of improving educational practice nationally, cooperation with personnel in key organizations of the state and nation is essential. An internal information section and an external network involving the Center and other agencies were established, starting in 1964, and are being extended in order a) to transmit information directly to scholars and others through Center supported publications, conferences, etc.; b) to communicate information to personnel of selected school systems and school-related educational agencies regarding the outcomes of the Center's programs as they become ready for dissemination and school use; c) to distribute knowledge regarding cognitive learning and products to agencies that function as linking mechanisms with schools; and d) to receive from the other agencies in the network information that may be used to improve the Center's products and operations.

Reports of research and descriptions of developmental activities are disseminated through various publications. In addition, all instructional programs and procedures produced and tested by the Center may be observed in selected schools. The Center is working with other agencies to assure statewide and national dissemination and adoption of Center products. The State Department of Public Instruction of Wisconsin, regional educational laboratories, Title I and Title III programs of ESEA, local schools of Wisconsin and other states, agencies of The University of Wisconsin, and the private sector are all involved in the dissemination network.

The Center's publication program carried out by the Office of Information Services has four major components – Technical Reports, Theoretical Papers, Working Papers, a Newsletter, and books. Technical reports describe the rationale, methods, and results of original research conducted by Center staff. Research reviews, statements of theoretical position, and descriptions and rationale for project development are the content of theoretical papers. The theoretical papers are formal documents that, like technical reports, may be classified as monographs. Practical papers contain information of immediate significance to educational practitioners in establishing or extending instructional programs – guidelines for beginning new programs, outlines and specific procedures for instruction, evaluation summaries and procedures, and progress reports of projects conducted in the schools. Working papers present plans and procedures for initiating and executing research and development activities.

Analyses of Concept Learning edited by Herbert J. Klausmeier and Chester W. Harris and the present book are collections of papers read and discussed at the Center. The first, from a conference of the same title, was published by Academic Press in the fall of 1966. The second, from the Center colloquium series, was published by Dembar Educational Research Services in 1968.

TRAINING ACTIVITIES

Three conditions indicate the need for a high priority to be given to the training of personnel in educational research and development in the R and D Center. First, the nation needs a large number of able persons in educational research and development. The many new programs under the ESEA of 1965 and others originating in the various states and local school systems require research and development personnel as well as knowledge and solutions to many educational problems. Second, The University of Wisconsin has both a tradition and the resources to contribute to the training of educational researchers, particularly beyond the baccalaureate. The University has strong graduate programs through which students acquire knowledge and skills essential for educational research. Third, the R and D Center at Wisconsin is the largest agency in the Midwest with educational research and development as its primary function.

The informal training of educational researchers and developers to supplement the formal programs in the various instructional departments and colleges, has been a significant contribution of the Center to improving education through research and development. For example, during 1966-67, informal training was provided for four postdoctoral fellows, sixty-three graduate students employed part time by the Center, ten ESEA Title IV doctoral fellows, ten NSF fellows, five NDEA fellows, several University fellows, and forty-two learning specialists employed by the public schools. Twenty-eight persons affiliated with the Center were awarded PhD's by the end of 1967 and are located throughout the country in a variety of leadership positions. A large group of persons has been awarded the Master's degree also.

TECHNICAL SUPPORT

The primary components of the Technical Section are a measurement and design group; production specialists; and computer experts. The design group provides assistance to the research personnel on matters of research design and evaluation. Equally important, this group is continuously developing new designs to meet the needs of the Center and refining statistical procedures for the analysis of data. A measurement expert provides assistance on problems of measurement and develops more efficient measurement tools and procedures. Computer specialists coordinate data analysis, from the conversion of raw data to computer compatible form through the presentation of the processed data in tables and graphs, and assist in the application of computer technology to educational research and development. A curriculum specialist assists the staff in identifying and using appropriate reference materials and in the production of graphical materials. Production experts assist with the development of instructional, experimental, and informative materials in the various media.

STAFF

Execution of the Center programs requires proper interdisciplinary staffing. This Center subscribes to the philosophy that the most productive research and the development of effective instructional systems will derive from teams composed primarily of behavioral scientists, subject-matter specialists, and experts in curriculum and other areas closely related to educational practice. The behavioral scientist contributes his research skills and knowledge of fundamental processes in cognitive learning; the subject-matter specialist provides knowledge of the subject field; the experts in curriculum and other fields contribute knowledge about methodology and other relevant instructional variables.

Studies of cognitive learning, included in Program I, are headed by behavioral scientists concerned primarily with synthesizing and advancing knowledge in this area. The other experts contribute to these studies as team members when cognitive processes related to specific subject content are studied. In Program 2 in which processes of instruction are studied and instructional programs are developed, the projects are headed by subject-matter or curriculum specialists concerned with developing an instructional system in a subject area; however, a behavioral scientist is a member of the team. Experts in school organization and the sociology of groups are involved with other staff in the development and maintenance of the facilitative environments, Program 3. Communication specialists carry out the dissemination program.

This mutually beneficial involvement of the staff has two natural consequences. First, the basic studies in learning are concerned with developing knowledge that is relevant to educational practices. Second, the studies of instructional processes and the instructional products that are developed in Program 2 have a sound foundation in the behavioral sciences and the subject disciplines.

In addition to the Center staff, many outstanding interdisciplinary groups are available within the University for cooperative planning and execution of activities. Personnel from three of the outstanding activities in The University of Wisconsin are continuously engaged in consultation and team work with the Center staff: WHA-TV, a foremost producer of educational TV programs in the nation; the Survey Research Laboratory, a large and distinguished interdisciplinary survey team; and the Computing Center, one of the best in the nation. Thus, the Wisconsin R and D Center does not require a large permanent staff in each of these areas of specialty. Rather, programs and activities are planned and executed cooperatively as part of the total University of Wisconsin contribution to improvement of education through research and development. Other relevant interdisciplinary centers and laboratories at the University include the Center for Demography and Ecology, M s Communications Research Center, and a Socia System Research Institute.

Twenty-two persons of professorial rank were affiliated with the Center, 1967-68. The departments in which they held rank included Computer Science, Curriculum and Instruction, Educational Administration, Educational Policy Studies, Educational Psychology, English, Psychology, Sociology, and Speech. Some professors held rank in two departments.

ADMINISTRATION

Administratively, the Center is located in the University's School of Education but draws upon the entire University for human and other resources. The working relationship between the Center and the University is manifested in three ways: professors in various University departments are members of the Center's Faculty; both the Dean of Education and the Chancellor of the University are involved in reviewing the Center's activities; and University officials serve on the Center's Executive Committee, the Policy Review Board, and the Advisory Council. Figure 4 outlines the administration and management of the Center.

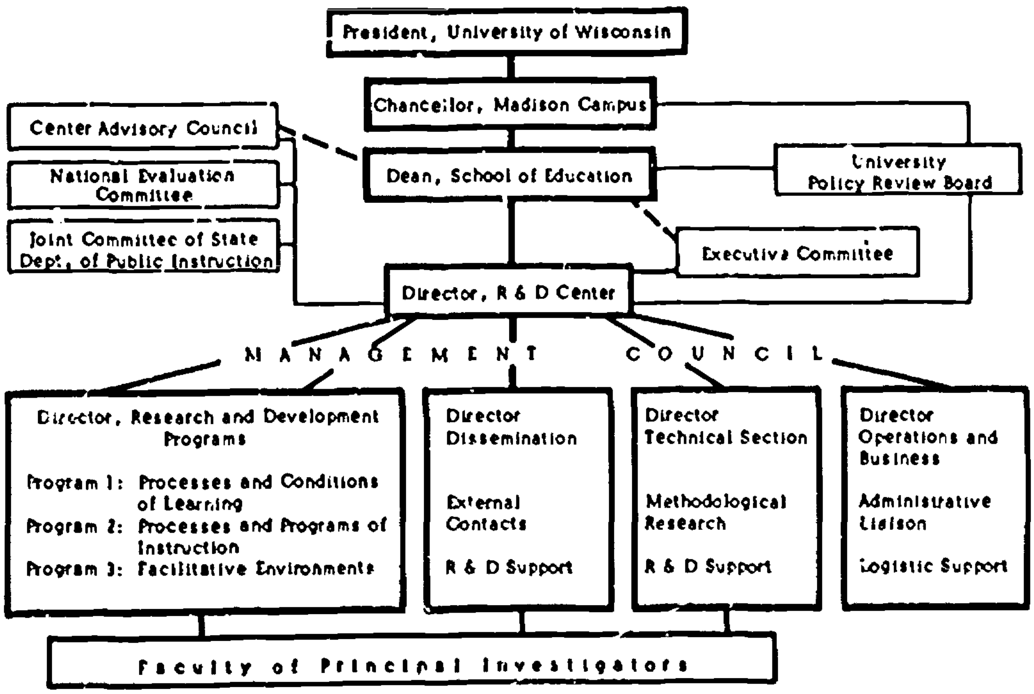
The Center Director is responsible for planning the total Center program and its broad strategies. A director of the research and development programs, a director of dissemination, a director of operations and business, and a director of a technical section, with the Center Director, comprise a Management Council.

The Executive Committee makes recommendations regarding policies, programs, and budget. It consists of the Center Director, who is also the chairman, three members elected by the Principal Investigators, and three members appointed by the Dean of the School of Education.

A Faculty of Principal Investigators, consisting of professors in various University departments, participate in the Center's program development, and they plan and execute the Center's projects.

The Policy Review Board brings the necessary University resources to bear upon the problems which the Center is interested in solving. The

Figure 4: Wisconsin Research and Development Center for Cognitive Learning organization chart



Board is composed of representatives from the office of the Chancellor of the Madison Campus; the deans of the School of Education and the College of Letters and Science; the Director of the Wisconsin Improvement Program; the Superintendent of State Department of Public Instruction; a member of the Wisconsin Coordinating Committee for Higher Education; the Executive Secretary of the Wisconsin Education Association; three departmental chairmen of the College of Letters and Science and the School of Education; and the Center Director. The Board annually reviews the Center's mission, organization, programs and staffing; recommends overall policies; suggests funding levels.

A National Evaluation Committee confers annually with Center personnel on the Center's progress. The Committee consists of nine individuals of national reputation interested in the problem area of the Center and committed to the improvement of education through research and development activities.

The Center Advisory Council, composed of officers of the University and of a representative of the Wisconsin School Board Association, meets once a year to analyze the Center's progress and to advise on matters in need of attention. Individuals of sixteen school systems, including public and parochial schools and a technical institute, are included.

LOCUS OF OPERATION

Because the R and D Center is effectively integrated into the University complex, both on an administrative and disciplinary level, it is natural that the Center program incorporates the philosophy of The University of Wisconsin regarding the improvement of society through research, development, and training. This philosophy is best expressed in a widely quoted statement:

> Whatever may be the limitations which trammel inquiry elsewhere, we believe that the great state University of Wisconsin should ever encourage that continual and fearless sifting and winnowing by which alone the truth can be found. (Taken from a report of the Board of Regents in 1894.)

President Harrington has from the beginning implied the extension of these concepts to educational research through the R and D Center, and he continues to give generously of his time to assist the Center in achieving its goals, and to assure a climate in which the Center may operate more effectively.

The Madison Campus of The University of Wisconsin is the foremost research agency in the state and is among the leading research-oriented universities in the world. The commitment to educational

research is substantial as may be inferred from a statement by then Chancellor Robben Fleming on July 1, 1965:

> With such emphasis on education in today's world it would be easy to drift along the well-marked paths of the past. The strength of our School of Education lies in its restless innovation. In cooperation with the schools of the state, it is forever trying new ways to improve the primary and secondary schools. In residence its curriculum is marked by an unusual degree of cross-fertilization, and by an interdisciplinary faculty. Its research is probing the frontiers of the learning process, asking, in effect, whether we are on the right path to begin with.

Implementation of the philosophy implied in the preceding statements as applied to the R and D Center has been demonstrated in two notable ways. First, the Policy Review Board of the Center, the most powerful governing agency of the R and D Center, is chaired by James Cleary, Vice-Chancellor of Academic Affairs. This Board reviews the program of the Center and takes major responsibility for assuring adequate staffing of the Center in light of its program. The second example of implementation is that the actual state support, not an indirect contribution, to the Center has increased from approximately $10,000 in 1965-66 to $108,000 in 1967-68. This state contribution is used almost totally to pay the salaries of principal investigators of the Center - professors who hold appointments in an academic department of the University and are budgeted fifty percent or more in the Center.

The School of Education is committed to the philosophy that the continuous improvement of public school education requires unceasing research and related activities. Dean McCarty, shortly after arrival at the University in 1966, indicated that the R and D concept is the most significant educational innovation in recent decades and that all forms of worthy educational research merit support. Three of many possible examples point to the high priorities given to educational research and related activities at The University of Wisconsin. First, $4,024,298 was expended on educational research and program experimentation in 1965 through the School of Education. The four sources of the support were Federal agencies, the State of Wisconsin, private organizations, and local school systems. Second, in the first seven years of the Cooperative Research Program of USOE, the University was awarded grants for thirty-three projects. Third, the amount of state support for research to the School of Education increased from $294,963 in 1964-65 to $804,976 in 1966-67.

The interdisciplinary nature of the School of Education is a major factor in the eminence of the University in educational research and related activities. In 1965 the membership in the School of Education was 1,075 professors from seventy-two departments of the University. Graduate work in preparing educational researchers is interdisciplinary. Professors in many departments of the University offer courses taken by graduate students pursuing programs to prepare for educational research. Further, many PhD candidates in the various departments of the School of Education pursue minors in departments of the College of Letters and Science while many PhD candidates in Letters and Science pursue minors in departments of the School of Education.

SOURCE AND STABILITY OF FUNDING

Any organization requires stable funding over a period of years in order to secure excellent staff at reasonable salary levels. Stability of funding was an important consideration in starting the Center. This Center, in line with the general USOE policy, was awarded a 5-year grant, starting September 1, 1964, by the USOE with three important provisions: First, that the contract would be amended annually; second, that, if for any reason, it became necessary to discontinue the Center the phasing out process would be accomplished over a period of time mutually satisfactory to The University of Wisconsin and USOE; and third, in the fourth year, the achievements and operations of the Center would be evaluated comprehensively in order to arrive at a decision regarding a second 5-year period. The Center was awarded approximately $500,000 for the first year of operation. Each amended contract for the next three years showed a modest increase in the level of support.

The second source of support is the State of Wisconsin through The University of Wisconsin. As noted earlier, some monies are made directly available to the Center. This is, however, not the main University support. The largest support lies in the availability of personnel to the Center - administrators, professors, graduate students, and classified personnel - and also other personnel in specialized fields such as computer science, opinion survey, TV production, and equipment development. The University also has leased a facility and provides the many services that are necessary to make an operation function well.

Although not a direct source of funds to support Center activities, the staff of four local school systems, particularly the subject-matter consultants, the building principals, and unit leaders, give considerable time to refining the concept of individually guided education in multiunit schools. Also, ideas and innovations of the local schools and the R and D Center are demonstrated through Title III projects in local schools. The State Department of Public Instruction is the principal disseminator of R and D field-tested ideas and materials in Wisconsin.

CONCLUDING STATEMENT

The preceding description of the Wisconsin Research and Development Center for Cognitive Learning was written late in the third year of the Center's existence. It is anticipated that there will not be sharp changes during the next decade in any of the eight major components: problem area, outcomes sought, research and development programs and strategies, dissemination program and strategies, pattern of staffing, pattern of administration and management, locus of operations, and source and

stability of funding. Though sharp changes are not expected, continuous improvement in all of these areas is encouraged. Annually the Executive Committee of the Center evaluates each program and project and the related strategies. The National Evaluation Committee gives particular attention to the significance of the problems being attacked and the adequacy of the strategies being employed. The University Policy Review Board examines all elements of the Center activities but gives particular attention to proper staffing. Field visitors appointed by USOE officials, and USOE officials also evaluate all components of the Center. University officials as individuals, consultants invited by the Center staff, and school people provide evaluative information. The information from all sources is carefully analyzed by all the Center staff, but particularly by the Executive Committee, Management Council, and Director. Annually, the information is formally put into program plans for the ensuing year. Changes are also made as essential at any time during the year. When any element does not function properly or produce at a reasonably high level, it is modified. The Center now demonstrates higher productivity and more sophisticated strategies than a year ago. Its improvement mechanisms are designed to accelerate these trends.

FOOTNOTE

1. Miss Elizabeth Schwenn, Project Assistant, assisted in editing this chapter. The Executive Committee and Management Council of the Center contributed to the program description of the Center and the various faculty and their research and development staff are responsible for the results as reported herein.

An Output-Oriented Model of Research and Development and Its Relationship to Educational Improvement

Hendrik D. Gideonse[1]
United States Office of Education
Washington, D. C.

FOR SEVERAL months I have followed the pages of Science attending to the continuing discussions of research and science policy, generally concerning myself with the degree to which the policy questions which arise in educational research and development can, must, or ought to be considered as a subset of those which arise for science as a whole. Thus it has been with no small degree of interest that I have read the articles prepared by Philip Handler (1967), George Daniels (1967), Harvey Brooks (1967), Lee DuBridge (1967), and Michael Reagan (1967).

Coincidental to this continuing dialogue I have been for the past ? years engaged in a general effort to improve the quality, effectiveness, and impact of the total research effort in education. No small part of that time has been spent grappling in practice with many of the same kinds of issues as receive explication in the articles cited above. For some months I have been thinking about research and development in education trying to develop a model which would express the different functions within the total research effort, the various sources of initiative for these different kinds of activity, and the relationships among both the functions and the sources of initiative. My thinking has, of course, been in no small measure stimulated by recent debate about the ways in which schools, instruction, and education are likely to be improved most quickly and with the most substantial cumulative and lasting impact.

Reviewing the several statements cited above in terms of their general usefulness for spotlighting and clarifying analogous problems in the field of educational research revealed the particular value of Reagan's analysis as a summary of points of view already advanced and as an identification of the generally individualistic starting point of many of the theoreticians of science and research policy. The model presented in this article constitutes, I think, an instance of exactly that for which Reagan called (1967), namely, a discussion of research and development as seen and understood from the point of view of someone who sees himself as a sponsor or user of research and development to achieve certain broad purposes in a particular field of human activity. It was not developed directly in response to Reagan's suggestion; in fact, it was only after a colleague recalled his article with specific reference to the similarity of view on the research and development distinction that I realized how closely what I had formulated came to responding to Reagan's suggestion for more dialogue on the part of users of research to more adequately balance that which had been generated by the performers.

The Guba-Clark Model

Typically, the context for the debate about research and development for education is a discussion of the change process in education. Various models of change have been proposed. The one which seems to have the highest degree of currency in education at the moment places research, development, and dissemination in a linear arrangement beginning on the left with inquiry (research) and proceeding to the right through development, diffusion, and adoption (see Figure 1). In proposing the model Egon Guba

Figure 1. A classification schema of processes related to and necessary for change in education*

		Development		Diffusion		Adoption		
	Research	Invention	Design	Dissemination	Demonstration	Trial	Installation	Institutionalization
Objective	To advance knowledge	To formulate a new solution to an operating problem or to a class of operating problems, i.e., to innovate	To order and to systematize the components of the invented solution; to construct an innovation package for institutional use, i.e., to engineer	To create widespread awareness of the invention among practitioners, i.e., to inform	To afford an opportunity to examine and assess operating qualities of the invention, i.e., to build conviction	To build familiarity with the invention and provide a basis for assessing the quality, value, fit, and utility of the invention in a particular institution, i.e., to test	To fit the characteristics of the invention to the characteristics of the adopting institution, i.e., to operationalize	To assimilate the invention as an integral and accepted component of the system, i.e., to establish
Criteria	Validity (internal and external)	Face Validity (appropriateness) --- Estimated Viability --- Impact (relative contribution)	Institutional Feasibility --- Generalizability --- Performance	Intelligibility --- Fidelity --- Pervasiveness --- Impact (extent to which it affects key targets)	Credibility --- Convenience --- Evidential Assessment	Adaptability --- Feasibility --- Action	Effectiveness --- Efficiency	Continuity --- Valuation --- Support
Relation to Change	Provides basis for invention	Produces the invention	Engineers and packages the invention	Informs about the invention	Builds conviction about the invention	Tries out the invention in the context of a particular situation	Operationalizes the invention for use in a specific institution	Establishes the invention as a part of an ongoing program; converts it to a "non-innovation"

*From Guba and Clark, 196[illegible]

and David Clark (1965) called three caveats to attention. First, they noted that the model was constructed on logical grounds and that it was largely unsupported by empirical research. Second, they pointed out that it was not necessary for change to begin at the research or inquiry stage. Third, they noted that the model itself was a uni-dimensional analysis of change roles which are influenced by a multi-dimensional range of variables not entirely accommodated by the model's structure.

As a model of the change process, this particular schema is simple and logical. However, those of us who have worked intensely on problems of research policy in education see some shortcomings. The Guba-Clark model does not emphasize sufficiently within its structure that initiative for action of quite different kinds can take place at any point in their continuum. Because of its linear nature, and despite the second caveat mentioned above, the model unwittingly implies that innovations begin with the findings generated by fundamental research.

The Output Model

The purpose of developing an alternative model is to create a heuristic which a) illustrates the essential differences between research and development activities and b) shows how the two are - or can be - related to one another and to the operating educational system. Such a model ought to illustrate the different sources of initiative and motivation for beginning various activities. It should be able to show or imply the interplay among all the functions in the effort to improve instruction.

The model developed is based on the conviction that research, development, and school operations are different kinds of activities with quite different objectives or outputs. It indicates that initiatives for each kind of activity are the results of decisions based on quite different kinds of data and equally different kinds of internal and external needs. The model implies that while there may be a strong logical flow from the production of knowledge through the development of processes to their installation in operational settings, there may be just as strong a flow backwards as operational problems define development programs, which, in turn, reveal the need for certain basic information and theory.

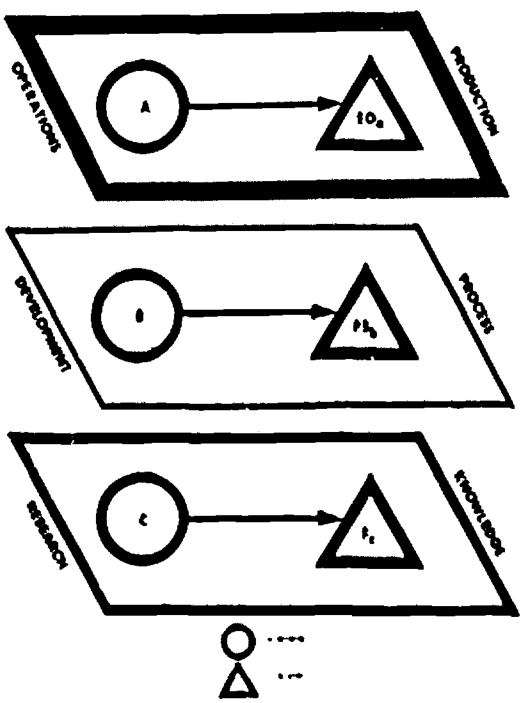
Figure 2 expresses the model. Three planes are shown, each symbolizing the different orientation of activities conducted under research, development, and school operations. (The model is - as are all models - an abstraction from reality. In the real world these activities are not always neatly separated either in time or location. The point of conceptually - and therefore graphically - separating them here is to illustrate the essentially different orientations of the three types of activity and the consequences of those differences.) For each activity represented in Figure 2, the model depicts an initiative leading to an output characteristic of that activity.

The lower plane symbolizes the knowledge orientation of research. The object of research is to generate new knowledge. One of the significant features of research is that when an activity is begun <u>the specific outcome is not known</u>. For research, C represents an initiative undertaken which culminates in a research finding represented by Fc.

The middle plane symbolizes what I call the process orientation of development. The object of development is to produce materials, techniques, processes, hardware, and organizational formats for instruction which accomplish certain pre-specified objectives construed to be part of the broader goals of education. One of the significant features of development is that when an activity is begun, <u>the objective is known or established at the outset</u>. The objectives for a development project are cast in the form of performance specifications (PS), and all activities are geared to producing the necessary products and processes which will meet those specifications. In Figure 2, B indicates an initiative undertaken for development culminating in the creation of a process which meets performance specifications PS_b.

The top plane symbolizes the activities characteristic of school operations. The operating educational system is production oriented. Thus, the object of school operations is to act upon human beings in order to train and develop in them various skills, attitudes, beliefs, and knowledge calculated to serve both society and themselves. Certainly one of the significant features of initiatives in school operations

Figure 2. An output model of educational research and development



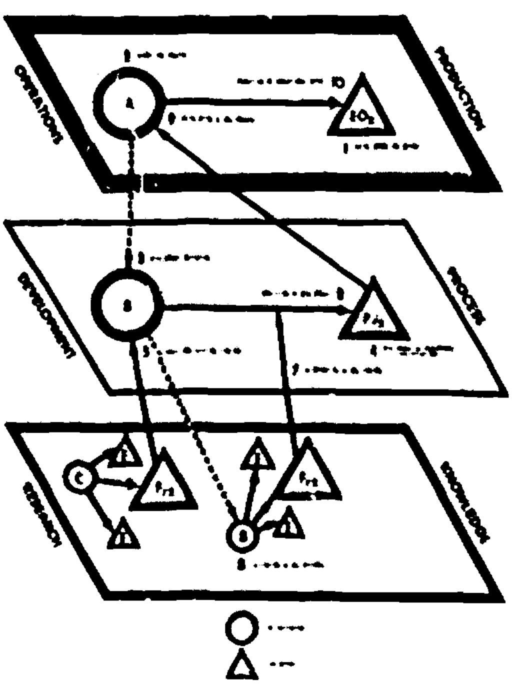
is the weight of the responsibility on the school administrator for choosing the right kinds of processes to achieve the outputs that society and individuals specify. In Figure 2, A represents an initiative to install a process leading to the production of education output EO_2.

A "Walk-Through" of the Model

To illustrate the relationships among the three types of activities, consider the following example. A responsible school official, faced with evidence that certain outputs desired by the society are not being achieved for a significant portion of the children in his charge, searches other school operations and ongoing or completed development projects for processes designed to meet his need. Should he find nothing to suit his particular problem (e.g., the low reading achievement of culturally disadvantaged children), he may then exercise his prerogative to call for the initiation of a project to design and develop a process whose performance specifications are such that upon installation of the process in his school, it will yield the desired educational output (e.g., increased level of reading achievement in the target population).

Once the initiative for the development project has been undertaken and the performance specifications established, the development project then conducts a search for relevant research findings which may offer clues to guide the development project.

Figure 3. A "Walk-Through" of the model



(Whether or not this step is taken after the project is begun or immediately before it is not really important. What is crucial is that at some point near the very beginning of the effort such a search is made.) Impressed with a particular finding (e.g., the tremendous impact of parental attitude on student achievement as revealed by the survey conducted as part of the Plowden study), the project may decide to develop a process which deliberately tries to engender a large measure of parental involvement in home instructional experiences carefully geared to complementary experiences in the school setting. Having made that decision the developers may then discover that they require further information about the specific nature of optimum parent-child interactions to stimulate maximum learner achievement. They may therefore call for a specific initiative of a research activity to generate further data to guide the development of materials. When useful findings are identified they can be incorporated in the development effort which then proceeds to successful conclusion. When, using iterative techniques of design, development, trial, and redesign on the basis of feedback, materials encompassing both school experiences and parent-child interactions in the home are successfully developed and validated, they may then be transferred to the operating setting where the administrator may install them as part of his instructional program.

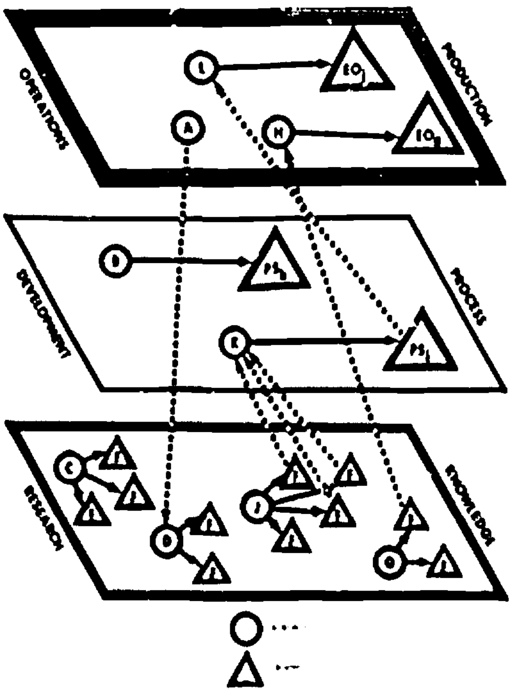
This example is illustrated in terms of the model by Figure 3. EO_x at No. 1 symbolizes the social demand for a certain kind of educational output (in the example just given, higher reading achievement for culturally disadvantaged children). This demand creates pressure on the school administrator to respond with some sort of initiative. That initiative is represented by A at No. 2. B symbolizes his search for an effective process to install. Since he did not find it, his response was to call for a development initiative (B at No. 3). The next step was to develop the performance specifications (PS_x at No. 4) such that they correspond to the educational output desired by society. Once the specifications for the development project are established, the next step is to survey related research seeking guidance for the development effort. The search is conducted and the finding (F_{rx}) of relevance to the performance specifications and the desired educational output (e.g., the significance of parental attitude) is incorporated into the development project (No. 5). The call for additional research assistance is symbolized by D at No. 6 and the incorporation of relevant findings (again F_{rx}) from that initiative into the development effort is symbolized by the solid arrow to the development line at No. 7. Number 8 represents the completion of the development project, No. 9 the incorporation of the process into school operations, and No. 10 the production of the desired output, higher levels of reading achievement, as a consequence.

A Sampling of Other Possible Interactions

The depiction of a sampling of other possible interactions among research, development, and operations can be found in Figure 4.

Example: A school official feels the need to assess the degree to which instructional programs are

Figure 4. A sampling of other possible interactions



serving a particular target population. He calls therefore for an initiative in research. This is represented by the D/D/F-F interaction.

Example: An organization engaged in development independently concludes that it would be useful to develop a certain process or product for instruction. This is represented by the B/PS$_b$ interaction.

Example: Research is initiated for its own sake and pursued solely for the knowledge which it produces. No findings have yet been incorporated either in development or operations. This is symbolized by the C/F-F-F relation in research.

Example: Research initiated for its own sake yields the findings that certain organizational structures for large city school systems are always problematical or that a certain vitamin supplement administered between the ages of 5 and 7 can prevent a form of mental retardation whose appearance cannot be detected until somewhat later. Neither one of these findings needs to pass through development. Each can be implemented directly in school operations (if so desired). This kind of relationship of research to practice is illustrated by the G/F-F/H/EO$_g$ interaction.

Example: Finally, consider an instance in the form of the linear flow or Guba-Clark model from research to development to implementation. Research on early childhood and cognitive growth, for example, uncovers a number of findings some of which may suggest the development of processes and environments which can actively enhance such growth. Development efforts are consequently supported and carried through to completion. The availability of the resulting products and processes is made known to educational administrators who are thereby persuaded to incorporate the newly developed early childhood instructional programs, at either private or public expense, into the Nation's educational system. This kind of sequence is represented by the J/F-F-F/K/PS$_j$/L/EO$_j$ interaction.[8]

THE MODEL AS AN HEURISTIC

All of these representations in Figures 2, 3, and 4 are fairly obvious and straightforward. The use of the model as an heuristic, however, profits from further explanation. One of these uses pertains directly to the problem of "change process" as applied to education. The model is structured to illustrate that the incorporation of research findings into development is just as important and oftentimes as difficult a proposition as incorporating newly developed processes into operational settings. The notion that there are obligations on both research and development to transfer their "products" to other activities means that each must pay careful attention to the way in which its outputs are presented and, perhaps the very way in which the outputs are produced. In other words, the requirement that eventually there be transfer or incorporation into another type of activity places constraints upon the professional behavior in each activity which cannot be ignored without compromising later impact.

This requirement is particularly true for development projects, but it is as true for research activities. A few concrete examples illustrate the point. One such example is the researcher who publishes his findings in a sloppy or difficult format and thereby hinders the likelihood of their being incorporated ultimately into practice. The researcher who inadvertently conceals or compromises his methodology or design encounters similar problems.

Similarly, the final requirement for a development project is that it be usable in operational settings. The ultimate desire to incorporate the developed process in school operations means that one of the performance specifications for development must always be the provision of procedures (teachers' manuals, training procedures, etc.) for accomplishing the installation of the innovations. If the development is undertaken without reference to that fact (if, in short, the requirement for transfer is not built into the performance specifications), the developer may well have rendered his product unusable. Hence, for example, the desire to involve teachers and other practitioners in the development process stems from the need to have their expertise and experience continually represented. They constitute, in effect, the embodiment of many of the operational constraints and possibilities in schools within which the finished process will operate. There are other good reasons for involving teachers as well, not the least of them being the respectability lent to the project in the eyes of the practitioners at large by

virtue of the meaningful presence of teachers in the effort. This last consideration is of no small importance in securing acceptance of the innovation in the profession at large. The nature of its importance, however, should not be mistaken; the involvement must be meaningful and not merely window-dressing, for the respectability is lent by their presence only if their contributions are fully utilized and their real knowledge of what is or may be possible is thoroughly weighed in the development effort.

One final point might be made about the model's portrayal of the possibility for transfers back and forth among research, development, and operations. That is the obvious emphasis which needs to be given to the problem of information flow and the need for carefully considering techniques for installation of better knowledge and better processes into their intended settings. Only part of this is the direct responsibility of the researchers and the developers. Those with obligations for considering the entire R and D system for improvement need also to direct their attention to the diffusion process.

A second feature of the model as an heuristic is the way it helps to clear up part of the problem of distinguishing between basic and applied research in education.[3] The model clearly implies that basic research (studies generated independently in research for the sake of the findings alone) and applied research (research conducted to serve a particular need identified by people engaged in development or operations) differ from one another primarily in terms of the intent of the initiator.[4] Thus the knowledge-orientation of the basic researcher is central to his activity. Applied research is also supported for the knowledge which results from it, but the initiator of the research knows to what instrumental use he is going to put his findings. By depicting both applied and basic research as similar kinds of activities, the model implies that <u>in and of themselves</u> they look very much alike. The procedures, the design, the sophistication must all be on par if either is to be valuable. What distinguishes the two from one another, if anything does, are the purposes for which they are initiated.[5]

A third feature of the model as an heuristic is its suggestion that decisions to initiate activities of each of the three types are made according to quite different criteria and perhaps by quite different people. The fundamental scientific character of research suggests that independent initiatives exercised there depend heavily on advice from the science community. Development projects, however, can also be independently initiated, but decisions to begin these kinds of activities are subject to advice from both research and operations. With limited resources, deciding which needs to satisfy through development (for example, those independently generated by developers, compared to those stemming directly from school operations, compared to those growing out of research activities) becomes a particularly difficult problem. Finally, the kinds of lonely decisions required of school administrators at the operational level are made by people in the context of still different circumstances and institutions. By emphasizing the essentially different nature of the activities being undertaken, the model reminds the policy maker of the need to collect different kinds of data and statements of need when planning future activities.

Finally, the frank attempt to represent each of the activities in terms of particular kinds of outputs may well be the most significant aspect of the model. It forces the viewer of the model to consider what the outputs of each activity are and to think about how the outputs of each activity are of use to one another. The outputs of research, for example, are knowledge. Some of the knowledge produced through research will find its way into development and into school operation. Are there ways of improving the output of research, making it more powerful, increase the likelihood of its being of use to instruction and education?

What about the outputs of development? They constitute, on the one hand, the validation of research and, on the other, the means by which the educational system can improve the manner in which it carries out its functions. How can development be improved, how can research be organized to be of greater use to development, and how can the educational system itself orient its organization to the recurring need for the installation of more powerful validated techniques?

Finally, what happens to educational operations when they begin to view their responsibilities in terms of output? The contrast can perhaps be most sharply drawn by considering the implications of the notion of grading <u>schools</u> on the basis of their <u>outputs</u> rather than <u>students</u> on the basis of their <u>performance</u>. The existing practice of grading <u>students</u> assumes at bottom that the student is responsible for his learning and that his failure or success is a tribute or a consequence of factors intrinsic to him. The idea of grading a <u>school</u> on the basis of its outputs assumes quite to the contrary that all students can learn and that the responsibility of the schools is to make that happen. (We do not, for example, judge the success or failure of medicine or law by the patient's or client's end state; we judge it by the degree to which the doctors or lawyers skillfully utilized the most sophisticated practices in attempting to serve the client. We certainly do not "grade" the patient or client; quite to the contrary, it is the professional services themselves which are assessed. An output orientation for school operations would cause the same reorientation of the direction of assessment in education.) If the schools themselves are judged in terms of the degree to which they are accomplishing their "production goals," increasingly they may come to orient their activities to assessing their own effectiveness, identifying the techniques and processes which need improvement and, as a consequence, calling with increasing sophistication for the kind of development activity and research support which will provide the basis for continuing improvement.

FOOTNOTES

1. The author is Director of the Program Planning and Evaluation Staff of the Bureau of Research, U.S. Office of Education. This paper is written in Gideonse' private capacity. No official

support or endorsement by the U.S. Office of Education is intended or should be inferred.

2. This example is, of course, clearly an ideal conceptualization. It is instructive to keep the model in mind as one looks back over the past four years at the tremendous developing interest in the establishment of *early childhood* educational programs. While it is certainly difficult if not impossible to establish an individual cause for such a complex phenomenon, it is nonetheless significant, I think, that Benjamin Bloom published his volume Stability and Change in Human Characteristics (1964) just at the time the Office of Economic Opportunity was beginning its planning toward the development of programs to fight the war against poverty. The significant thing about Bloom's volume, however, was that while his conclusions firmly underscored the importance of the early years in the development of cognitive skills, he also pointedly observed that there had been little actual experimentation designed to create environments to enhance such skill development. I do not think that it does violence to Bloom's argument to interpret it as a call for rigorous development efforts designed to produce environments and instructional programs which have the effect of enhancing human capabilities. The problem, however, is that which exists in all social domains. There are large numbers of children now whom we cannot afford to ignore, and the consequence has been the attempt to create operating early childhood programs based on those convincing research findings without first having gone through a developmental stage. The result has been a somewhat marginal impact on the target population despite the clear implications of the research summarized by Bloom.

3. I think it is also useful at this point to recall Reagan's view that the social sciences present something of a different picture in regard to the distinction between basic and applied research. As he points out (Reagan, 1967, p. 1385), no matter how abstruse and abstract its practitioners attempt to become, social science research is inherently related to potential applications.

4. The initiator and performer are, of course, not necessarily one and the same person or institution. The actual initiator of the research project may be a school man, a developer, a researcher, or a research administrator. The performer of that research effort may or may not have the same ultimate purpose as the initiator in mind as he undertakes the activity. Thus, for example, it would be perfectly possible for a research administrator to stimulate a series of research activities relating to reading which he views as applied research necessary for a development effort to build improved reading curricula, while the performer of that research sees it as a basic research effort in perception or the psychomotor skills associated with reading.

5. This view, I believe, fits fairly well with one part of Harvey Brooks' analysis of the distinctions between basic and applied research (Brooks, 1967, p. 1706) when he noted that "as definite categories, (the terms) basic and applied tend to be meaningless, but as positions on a scale within a given environment they probably do have some significance." The principal shortcoming with Brook's analysis, however, is that it proceeds almost entirely on the presumption that the distinction can be resolved by approaching it in some way from the researcher's point of view. My experience in the administration of research, as the model presented in this paper clearly indicates, leads me to believe that the researcher's view is only one of several which must be taken into account in attempting any analysis of the distinction between basic and applied research. The presence of rather different criteria and alternative vantage points convinces me that Reagan's argument for abandoning the distinction within research is much more persuasive as is the suggestion that the critical categorization is that between research and development.

REFERENCES

Bloom, B. S. Stability and change in human characteristics. New York: Wiley, 1964.

Brooks, H. Applied science and technological progress. Science, 1967, 156, 1706-1712.

Daniels, G. H. The pure-science ideal and democratic culture. Science, 1967, 156, 1699-1705.

DuBridge, L. A. University basic research. Science, 1967, 157, 648-650.

Guba, E. G., and Clark, D. L. An examination of potential change roles in education. Paper read at Symposium on Innovation in Planning School Curricula held by National Education Association Committee for Study of Instruction, Airlielhouse, Virginia, October, 1965.

Handler, P. Academic science and the federal government. Science, 1967, 157, 1140-1146.

Reagan, M. D. Basic and applied research: A meaningful distinction? Science, 1967, 155, 1383-1386.